BioOne COMPLETE

Fitting and comparing models of phyletic evolution: random walks and beyond

Author: Hunt, Gene

Source: Paleobiology, 32(4) : 578-601

Published By: The Paleontological Society

URL: https://doi.org/10.1666/05070.1

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

Paleobiology, 32(4), 2006, pp. 578–601

Fitting and comparing models of phyletic evolution: random walks and beyond

Gene Hunt

Abstract.—For almost 30 years, paleontologists have analyzed evolutionary sequences in terms of simple null models, most commonly random walks. Despite this long history, there has been little discussion of how model parameters may be estimated from real paleontological data. In this paper, I outline a likelihood-based framework for fitting and comparing models of phyletic evolution. Because of its usefulness and historical importance, I focus on a general form of the random walk model. The long-term dynamics of this model depend on just two parameters: the mean (μ_{step}) and variance (σ^2_{step}) of the distribution of evolutionary transitions (or ''steps''). The value of μ_{step} determines the directionality of a sequence, and σ^2_{step} governs its volatility. Simulations show that these two parameters can be inferred reliably from paleontological data regardless of how completely the evolving lineage is sampled.

In addition to random walk models, suitable modification of the likelihood function permits consideration of a wide range of alternative evolutionary models. Candidate evolutionary models may be compared on equal footing using information statistics such as the Akaike Information Criterion (AIC). Two extensions to this method are developed: modeling stasis as an evolutionary mode, and assessing the homogeneity of dynamics across multiple evolutionary sequences. Within this framework, I reanalyze two well-known published data sets: tooth measurements from the Eocene mammal *Cantius*, and shell shape in the planktonic foraminifera *Contusotruncana*. These analyses support previous interpretations about evolutionary mode in size and shape variables in *Cantius*, and confirm the significantly directional nature of shell shape evolution in *Contusotruncana*. In addition, this model-fitting approach leads to a further insight about the geographic structure of evolutionary change in this foraminiferan lineage.

Gene Hunt. Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20013-7012. E-mail: hunte@si.edu

Accepted: 7 June 2006

Introduction

Phyletic evolution is most directly observed from repeated samples of populations from the same lineage traced through a stratigraphic section. When correctly inferred, these sequences capture evolutionary patterns and avoid the uncertainties associated with unobserved ancestral phenotypes (Cunningham et al. 1998). Paleontologists have long documented such evolutionary sequences, and these efforts have produced some of the best known—and most hotly debated—paleontological case studies (e.g., Trueman 1922; Hallam 1968; Gingerich 1976; Gould and Eldredge 1977; Raup and Crick 1981; Malmgren et al. 1983; Gould 2002).

With the advent of statistical approaches in paleontology, there came increasing effort to evaluate patterns in the fossil record in light of simple null models. This approach was thought to introduce rigor and protect paleontologists from ascribing complex causes to patterns that may actually be generated by a few simple rules (Raup and Gould 1974; Bookstein 1987). For evolutionary sequences, Raup (1977) introduced as a null model the unbiased random walk, a simple model of evolutionary change in which trait increases and decreases were equiprobable. In subsequent applications and extensions (Raup and Crick 1981; Bookstein 1987, 1988), workers often cautioned that apparent trends in evolutionary series should be interpreted only if the unbiased random walk model can be rejected. However, the low statistical power of these tests may cause this strategy to be overly conservative, especially when applied to time-averaged and incomplete paleontological sequences (Roopnarine et al. 1999; Sheets and Mitchell 2001b).

More recently, several authors have used random walks as part of a broader effort to characterize the patterns of phenotypic evolution, and in particular distinguish random

© 2006 The Paleontological Society. All rights reserved. 0094-8373/06/3204-0004/$1.00

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

walk, stasis, and directional modes of change (Gingerich 1993; Roopnarine 2001). These methods represent a subtle shift in the role of random walk models, relative to previous work. Rather than a null model to be rejected, a random walk is instead evaluated as one of several possible evolutionary modes, with the goal of choosing the mode that is most consistent with the observed sequence of morphological changes. In this paper, I describe a methodology that extends this strategy further. Under this approach, different notions about evolutionary mode are formalized into statistical models, and these models are evaluated on the basis of their likelihood of producing the observed data. The advantage of this framework is that it simultaneously allows for flexibility and statistical rigor. The flexibility follows from the great variety of evolutionary models that may be considered, as long as they can be represented by a suitable likelihood function. The rigor is a consequence of adopting a likelihood framework, with its well-established means for fitting and comparing models.

I begin with the simple random walk model that serves as the starting point to this approach, and proceed to describe how the parameters of this model may be estimated from paleontological data using maximum likelihood. I then present simulations demonstrating that the estimation procedure performs well, even when evolutionary sequences are very incompletely sampled. Next, I consider two extensions of the method beyond a simple random walk. The first extension provides a model for stasis in this framework, and the second considers the analysis of multiple evolutionary sequences. Then, I discuss the procedures available for choosing among several candidate models in the likelihood framework, with an emphasis on methods that do not require one model to hold privileged null status. Finally, I illustrate this general approach by reanalyzing two well-known published data sets (Clyde and Gingerich 1994; Kucera and Malmgren 1998).

General Random Walk Models

The starting point for this approach is the general random walk model as described by Hunt (2004). Time in this model consists of discrete intervals, during each of which an increment of evolutionary change is drawn at random from a distribution of evolutionary ''steps'' (Fig. 1). This distribution of evolutionary steps can be of any form, continuous or discrete, but it is assumed that this distribution does not change over time. For reasons discussed below, the long-term dynamics of a population evolving according to this model are governed almost entirely by the mean (μ_{step}, the mean step) and variance (σ^2_{step}, the step variance) of the distribution of evolutionary steps. The mean of the step distribution determines the inherent directionality of the sequence. On average, trait values will increase over time when μ_{step} is positive (Fig. 1B), decrease when μ_{step} is negative, and show no net trend when μ_{step} is equal to zero (Fig. 1A). In this last situation ($\mu_{step} = 0$), the sequence is said to be an unbiased or symmetric random walk; all other values of μ_{step} result in biased random walks. The variance of the step distribution (σ^2_{step}) governs the volatility of an evolutionary sequence. When σ^2_{step} is high, evolving populations will experience large increases and large decreases in trait values and accordingly volatile trajectories (Fig. 1C). In contrast, when σ^2_{step} is small, evolutionary steps will be consistently close to μ_{step} and evolutionary trajectories will proceed at a steady pace. When σ^2_{step} is very small and μ_{step} is zero, lineages will exhibit little phenotypic change over time. In short, μ_{step} corresponds to directional trends in trait values, whereas σ^2_{step} measures the magnitude of fluctuations around such trends, if present.

Using this formulation of the random walk model, all information about directionality in an evolving lineage is contained in the mean of the step distribution (μ_{step}). This is a departure from most previous studies, which have modeled directionality using the probability that each evolutionary step exceeds zero (p). According to this parameterization, setting $p = 0.5$ is used to model unbiased random walks. However, this strategy works only when the step distribution is symmetric; otherwise, evolution will be directional even when $p = 0$. Because long-term directionality

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

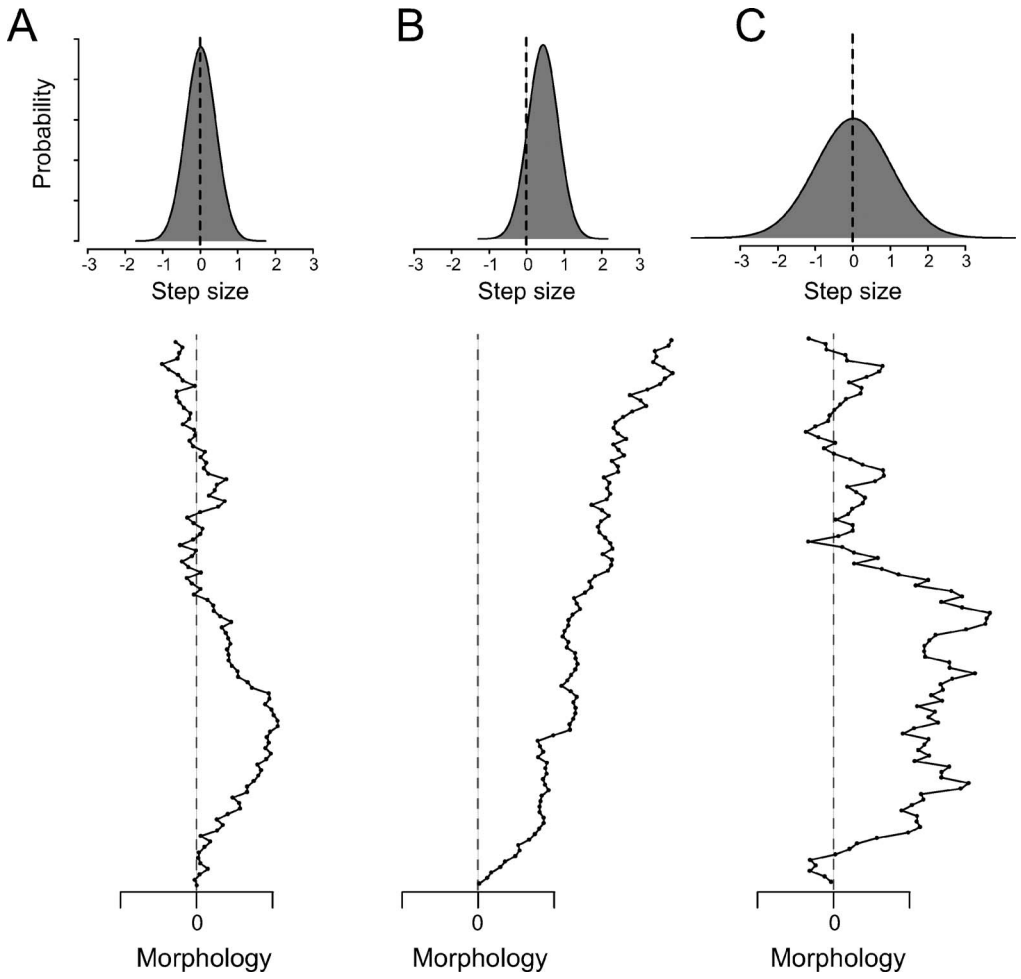

FIGURE 1. Three example step distributions (top) used to generate corresponding evolutionary sequences of 100 steps (bottom). When the mean of the step distribution is zero, increases and decreases are equally likely and the overall dynamics are nondirectional (A, C). Step distribution B has a positive mean, and therefore will tend to produce positively trended evolutionary sequences. With increasing step variance, evolutionary sequences are more volatile, with larger positive and negative excursions (compare C with A).

depends only on μ_{step} (see below), it is a natural measure of directionality.

Before proceeding, it is necessary to clarify some terminology. What I refer to here as a general random walk includes the whole class of models that are characterized by evolutionary transitions that (1) are independent from each other and (2) are homogenous over time; i.e., they are drawn from the same step distribution through the interval of interest. When unqualified in the paleontological literature, the term ''random walk'' generally implies an unbiased random walk, and some would restrict the term only to this subset of models. Throughout this paper, I use ''general random walk'' to refer to all models that meet the above two criteria and ''unbiased random walk'' to denote the special case of nondirectional random walks ($\mu_{step} = 0$).

It is important to note that modeling phyletic evolution as a random walk does not imply any particular evolutionary process; many different microevolutionary scenarios produce evolutionary sequences that can be described as random walks (Hansen and Martins 1996; Roopnarine et al. 1999). Although phenotypic transitions are modeled as random draws from a step distribution, this should not be understood to imply that the evolutionary changes themselves are unrelat-

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

ed to deterministic factors such as natural selection. Rather, randomness here is a consequence of causal ignorance (Raup 1977). Because we usually do not know the detailed mechanisms mediating each phenotypic change in the past, we limit ourselves to the more modest but attainable goal of inferring something about the aggregate qualities of a set of evolutionary changes, i.e., their directionality and volatility.

Statistical Inference and Random Walks

Although many paleontological studies have simulated evolutionary sequences as random walks, the inverse problem of estimating the generating parameters of a random walk from real evolutionary sequences has received little attention. In this section, I outline how to infer the parameters of the general random walk using likelihood. I first consider the expected distribution of trait differences between ancestral and descendant fossil populations, and then incorporate the contribution of sampling error to this expected difference between two population means. These two components allow for the calculation of the likelihood of specific values of μ_{step} and σ^2_{step} from a set of ancestor-descendant evolutionary transitions. Finally, I discuss the exploration of the parameter space to discover values of μ_{step} and σ^2_{step} that are most likely to have produced empirical data, along with some of the properties of these maximum-likelihood estimates.

Before getting to the mechanics of how to fit random walk models to paleontological data, it is worth considering how one could test if an observed evolutionary sequence is consistent with a random walk model at all. The relevant assumptions are that evolutionary steps are independent and homogenous over time. One way to assess this is to fit a random walk model along with a variety of other models representing alternative evolutionary dynamics, and determine which model (or models) is most consistent with the observed data. I discuss this general approach in some detail later in the paper. Testing for the homogeneity of the step distribution over time, in particular, is probably best done in this model-comparison framework. The independence of steps can be evaluated even in the absence of alternative models by testing for significant temporal autocorrelation among observed evolutionary transitions. Such tests can be based on contingency tables testing whether trait increases are more likely to follow increases or decreases in previous steps. Standard time-series tests for autocorrelation may also be used, bearing in mind that (1) the differences between adjacent sampled populations, not the time series of actual trait values, should be tested for autocorrelation, and (2) the shared error term between adjacent evolutionary transitions causes a negative autocovariance, which should be accounted for in assessing autocorrelation (the expected magnitude of this negative autocovariance is the sampling variance of the trait mean [see Bookstein 1987]).

The Distribution of Ancestor-Descendant Trait Differences.—Consider a morphological trait X evolving according to the general random walk model. At the start of the sequence, the ancestral population has a mean trait value of X_A and after t time steps, the final descendant population has a mean trait value of X_D (in this context, the ancestor and descendant populations represent older and younger samples, respectively, of the same lineage). For now, assume that these population means are known without error. The change in morphology ($\Delta X = X_D - X_A$) is simply the sum of all t intervening evolutionary steps (s):

$$\Delta X = \sum_{i=1}^{t} s_i. \quad (1)$$

Mathematically, ΔX is a sum of independent random variables. The evolutionary steps (s_i) are random in the statistical sense of being drawn from a probability distribution, and their independence is a basic assumption of the model. As the sum of independent variates, ΔX has a mean and variance that are equal to the sum of the means and variances (respectively) of its component variables (Bulmer 1979). The mean and variance of each evolutionary step are μ_{step} and σ^2_{step}, and so their sums over t time steps are $t\mu_{step}$ and $t\sigma^2_{step}$. Moreover, the Central Limit Theorem holds that the sum of independent variables increas-

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

ingly resembles a normal distribution as the number of variables increases. Consequently, as long as many time steps separate ancestral and descendant populations, the trait differences between populations separated by t time steps will be approximately normally distributed with a mean of $t\mu_{step}$ and a variance of $t\sigma^2_{step}$.

For paleontologists, this tendency for evolutionary differences to be normally distributed is a mixed blessing. Ideally, we would like to know the shape of the step distribution because it presumably reflects the processes generating evolutionary change. There has been much interest in knowing, for example, if the distribution of evolutionary changes is smoothly unimodal or if it has separate modes for stasis and punctuation. However, as the temporal window of observation expands, the distribution of evolutionary differences will become increasingly Gaussian, regardless of the underlying step distribution. As consolation for this information loss, there is a great practical benefit: for a lineage evolving according to a general random walk, all the information relevant for long-term evolutionary dynamics is contained in just the mean and variance of the step distribution. This greatly simplifies the problem of statistical inference by reducing to two the number of quantities to be estimated.

Incorporating Sampling Error.—The foregoing section assumed that trait means are known without error, or at least that this error is small enough to be ignored. Paleontologists are rarely so fortunate. More typically, specimens are scarce and evolutionary changes are subtle, and consequently sampling error contributes substantially to observed differences among samples.

We can account for sampling error by decomposing observed trait means into two components: the true population means (X_A and X_D), and error terms (e_A and e_D) representing the deviation of the observed mean from the true mean due to sampling error (Fig. 2). Including these error terms, the observed difference between ancestral and descendant trait means is

$$\Delta X = (X_D + e_D) - (X_A + e_A).$$

Substituting the sum of all evolutionary steps (s_i) for the difference in the true mean ($X_D - X_A$) gives

$$\Delta X = \sum_{i=1}^{t} s_i + e_D - e_A. \quad (2)$$

The estimation errors around the mean (e_D and e_A) are normally distributed with a mean of zero and variance V_p/n (Sokal and Rohlf 1995), where V_p is the population phenotypic variance and n is the number of measured individuals. The observed difference between ancestor and descendant is therefore a sum of three independent variables: Σs_i, e_D, and $-e_A$. Because these variables are normally distributed (or nearly so), their sum is also normally distributed (Bulmer 1979). Consequently, ΔX is normally distributed with a mean of $t\mu_{step}$ and variance equal to

$$\mathrm{Var}(\Delta X) = t\sigma^2_{step} + V_{p_D}/n_D + V_{p_A}/n_A, \quad (3)$$

which is the sum of the variances of its three component variables.

Likelihood of Ancestor-Descendant Transitions.—The results of the previous section define the probability of observing a particular evolutionary difference (ΔX) between populations separated by t time steps: ΔX is normally distributed with a mean of $t\mu_{step}$ and variance given by equation (3). The probability density function for a generic normal variate (y) is

$$\Pr(y) = \frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{[y - M]^2}{2V}\right), \quad (4)$$

where M and V are the mean and variance of the normal distribution. For evolutionary differences, $M = t\mu_{step}$ and V is given by equation (3):

$$\Pr(\Delta X) = \frac{1}{\sqrt{2\pi\ \mathrm{Var}(\Delta X)}} \times \exp\left(-\frac{[\Delta X - t\mu_{step}]^2}{2\ \mathrm{Var}(\Delta X)}\right). \quad (5)$$

Equation (5) gives the probability of observing a particular value of ΔX, given a hypothesis of specific values of μ_{step} and σ^2_{step}. This quantity is proportional to the likelihood of μ_{step} and σ^2_{step}, given the observed data ΔX. It is convenient to work with the natural logarithm of the likelihood, called the log-likelihood, or support (Edwards 1992). Taking the natural

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

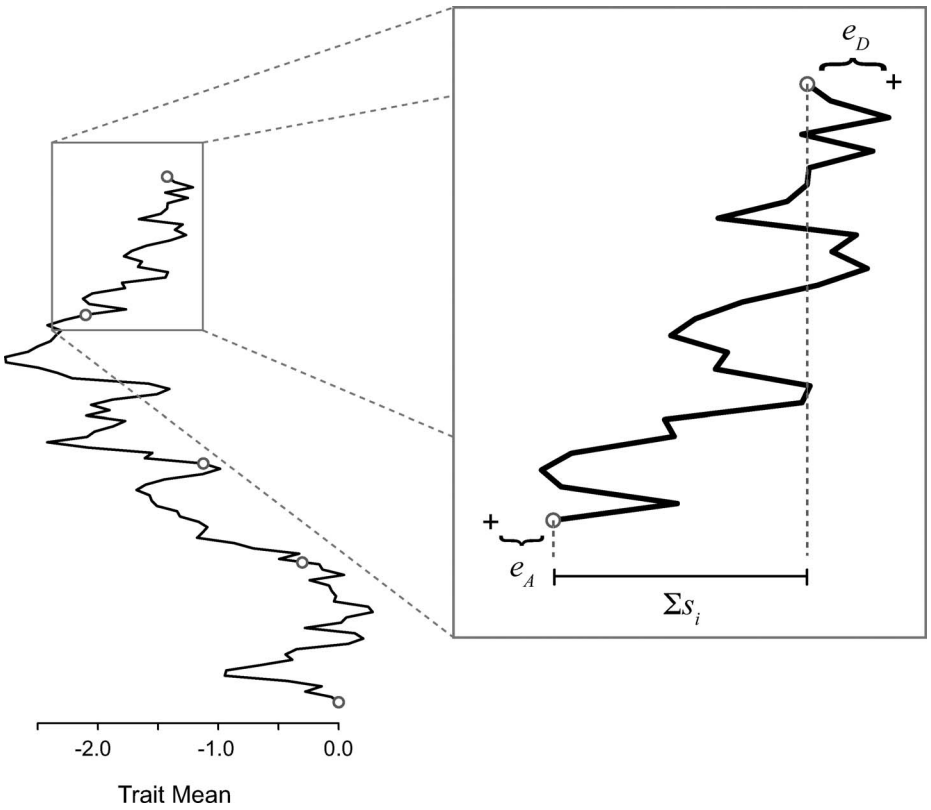

FIGURE 2. Mean trait in an evolving lineage (black line), sampled at five points in time (gray open circles). Close-up shows trait evolution between the last two sampled populations. The true trait difference between these two samples is equal to the sum of all evolutionary steps (s_i) separating the sampled populations. Because of sampling error, the estimated trait means (+) will differ from the true means by an error term (e). The observed difference between two populations includes both evolutionary differences (Σs_i) and sampling error (e).

logarithm of equation (5) and substituting equation (3) for Var(ΔX) yields the log-likelihood function of a hypothesized μ_{step} and σ^2_{step} for an observed evolutionary change ΔX:

$$\ell(\mu_{step}, \sigma^2_{step}) = C - \frac{1}{2}\log\left(t\sigma^2_{step} + \frac{V_{pA}}{n_A} + \frac{V_{pD}}{n_D}\right) - \frac{(\Delta X - t\mu_{step})^2}{2(t\sigma^2_{step} + V_{pA}/n_A + V_{pD}/n_D)}, \quad (6)$$

where C is a constant equal to $-\frac{1}{2}\log(2\pi)$, and all logs are natural logarithms. Equation (6) provides a measure of the relative support for rival hypotheses about the parameters of a random walk; hypothesized values of μ_{step} and σ^2_{step} are favored to the degree they produce high log-likelihoods according to this equation. The optimal estimates of μ_{step} and σ^2_{step} are those that maximize the log-likelihood of observed ancestor to descendant differences. Multiple ancestor-descendant trait differences may be used jointly to estimate μ_{step} and σ^2_{step} by summing the log-likelihoods (eq. 6) over all observed evolutionary transitions (Fig. 2).

The summation of log-likelihoods across multiple evolutionary transitions assumes that the transitions are independent observations. With transitions arrayed in a single evolutionary sequence, however, a slight dependence is introduced between adjacent trait differences because they share a sample mean and its sampling error (e.g., in a sequence of three samples, both evolutionary transitions involve the middle sample). This weak dependence should not bias the estimates of μ_{step} and σ^2_{step}, and indeed the good statistical properties of the estimation procedure (see below) suggest that the effects of this dependence are minor.

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

In order for these equations to apply to real paleontological data, it is necessary to relate time in the model, which is composed of discrete steps at which evolutionary changes occur, to the actual time scale by which fossil ages are measured. In other words, do the evolutionary steps occur every year, every thousand years, or at some other frequency? This question is unanswerable, but fortunately unimportant. As long as consistency is maintained within an analysis, results are unaffected by how frequently (in real time) model steps occur. Altering the assumed duration of steps amounts to changing the units in which time is measured, which scales parameter estimates accordingly (e.g., parameter estimates are one thousand times larger when steps occur every Kyr rather than every year). Because overall statistical inference is otherwise unchanged by this scaling, researchers can just measure time in whatever units are convenient.

Finally, note that an unbiased random walk is just a special case of the general walk model in which μ_{step} is constrained to be zero. As a result, its log-likelihood function is the same as equation (6), except that μ_{step} is replaced by zero.

Implementing the Method.—Equation (6) defines a three-dimensional surface in which the μ_{step} and σ^2_{step} define the *x*- and *y*-axes, and the log-likelihood is the height of the surface above the *xy*-plane. The maximum-likelihood estimates for μ_{step} and σ^2_{step} are those values corresponding to the highest peak in this likelihood surface. For the special case of constant temporal spacing and equal sampling error in all samples, it is possible to derive simple expressions that yield maximum-likelihood estimates of μ_{step} and σ^2_{step}:

$$\hat{\mu}_{step} = \overline{\Delta X}/t, \tag{7}$$

$$\hat{\sigma}^2_{step} = \frac{1}{t}(\mathrm{Var}[\Delta X] - \varepsilon), \tag{8}$$

where ε is the sampling variance for each evolutionary transition ($\varepsilon = 2V_p/n$) and t is the interval of time separating each adjacent sample (Appendix). These estimators are intuitively quite reasonable: μ_{step} is the mean evolutionary transition divided by elapsed time, and σ^2_{step} is the variance of the evolutionary transitions with the contribution from sampling error subtracted, also divided by time. These estimators seem to be fairly robust to mild violations of the assumptions of evenly spaced and uniform sampling error, but can perform poorly if sample spacing and sample sizes vary substantially.

For the more general case of variably spaced samples with uneven sampling errors, equation (6) does not lend itself readily to analytical maximization. Instead, numerical hill-climbing algorithms must be used to locate the optimum of the log-likelihood surface. Unlike more parameter-rich models (e.g., Foote 2003), likelihood surfaces for random walk models are smooth and unimodal, and pose little challenge for numerical optimization. For a single ancestor-descendant transition when evolution is assumed to be nondirectional, σ^2_{step} maybe estimated more simply from the within- and among-group variance components, as done by Lynch (1990). However, combining information from multiple evolutionary transitions over varying temporal intervals requires parameters to be optimized numerically rather than analytically.

In equation (6), there are several pieces of information that are required to compute log-likelihoods besides μ_{step} and σ^2_{step}: the sample sizes (n_A, n_D) and phenotypic variances (V_{pA}, V_{pD}) of the ancestor and descendant populations, and the number of time steps (t) separating these populations. Whereas sample sizes are effectively error-free, the same will not be true of phenotypic variances and ages of sampled populations.

Error in estimating phenotypic variances (V_p) will generally have little effect on statistical inference of random walks. Because it is nondirectional, this error should not bias parameter estimates upward or downward, although it will contribute slightly to estimation error. When some or all samples in an evolutionary sequence are poorly represented, it may be beneficial to assume that V_p is constant over all populations. If the data are consistent with this assumption (as indicated by a test for variance heterogeneity [Sokal and Rohlf 1995]), one can then substitute each separate estimate of phenotypic variance with the var-

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

iance pooled over all samples, greatly increasing the precision with which V_p is estimated.

Age model uncertainty is potentially a more serious problem. Because of the limits of correlation and absolute dating, paleontological age estimates will sometimes be systematically inaccurate, and usually rather imprecise. Systematic dating error occurs when all estimated durations differ from their true values by some multiplicative factor c. In such cases, estimates of μ_{step} and σ^2_{step} will change by a factor of $1/c$ because time only enters the log-likelihood function (eq. 6) in a product with either μ_{step} or σ^2_{step}. If time is increased by a factor c, multiplying μ_{step} and σ^2_{step} by $1/c$ leaves the log-likelihood function unchanged. This same reasoning also explains why changing the units with which time is measured simply scales the parameter estimates but does not otherwise affect the results.

Non-systematic (random) dating error has a more variable influence on this method, with estimation performance differing somewhat depending on the magnitude and distribution of dating error, and on the evolutionary dynamics of the trait. This context dependence makes it difficult to explore exhaustively the possible consequences of dating error. In my own preliminary investigations, the effects of temporal uncertainty were generally minor. Estimation of μ_{step} seems to be unaffected by random error in temporal durations, except for a slight increase in estimation error. Estimation error for σ^2_{step} increases more markedly in the presence of random age model error, but its contribution to overall estimation error is still quite small. Under some circumstances, estimates of σ^2_{step} become biased upward (overestimates), but this effect appears to be substantial only when temporal uncertainty is very high and evolutionary sequences are unrealistically directional. Although this issue requires further exploration, it seems thus far that age model uncertainty will usually not unduly hinder statistical inference of μ_{step} and σ^2_{step}. It may be prudent, however, for users of this method to explore by simulation the likely effects of realistic age model errors in the sequences under study.

I have implemented the estimation of μ_{step} and σ^2_{step} using the statistical programming environment R, version 2.1.1 (R Development Core Team 2005), taking advantage of the built-in capabilities for multivariate optimization. R is freely available for all common computer operating systems (http://www.r-project.org/). Files with the R functions necessary to perform the analyses in this paper have been bundled into an R package called ''paleoTS'' that is available for download from the Comprehensive R Archive Network (http://cran.r-project.org/).

Parameter Estimation Performance

Overview

To evaluate the performance of the estimation procedure, I applied it to evolutionary sequences generated from random walk models with known values of μ_{step} and σ^2_{step}. I focus on estimation accuracy (whether, on average, the procedure correctly estimates true parameter values) and to a lesser extent on estimation variance (precision). There are many quantities that conceivably could be varied in testing performance, but I will present just a subset of the simulation results that are most relevant to estimation of μ_{step} and σ^2_{step} in practical situations. Those variables that have little influence on log-likelihoods or would be largely outside the control of the investigator (V_p, sample ages, number of individuals per sample) were held constant at plausible values.

After describing the simulation procedure and evaluating the method's performance for a variety of evolutionary dynamics corresponding to different values of μ_{step} and σ^2_{step}, I then consider explicitly the effects of sequence incompleteness (Roopnarine et al. 1999) on this method. Readers concerned more with the conclusions than with details of these simulations may wish to skip forward to the ''Method Performance Summary'' section.

Simulation Procedure

Trait means were modeled as a random walk, with each evolutionary step drawn from a normal distribution with mean μ_{step} and variance σ^2_{step}. To incorporate the effects of sampling error, at each step 30 individuals were sampled assuming that trait values were distributed normally with unit variance ($V_p = 1$).

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

Unless otherwise noted, all sequences were run for 1000 time steps and then subsampled every 50 steps to produce sequences of 20 evenly spaced evolutionary transitions. Values of σ^2_{step} were varied from 1% of phenotypic variance ($0.01V_p$) to ten times phenotypic variance ($10V_p$), and μ_{step} was varied from zero to 10% of phenotypic variance, in both positive and negative directions. Although realistic ranges for these parameters are not yet known, these values of μ_{step} and σ^2_{step} were chosen visually to approximate the behaviors observed in paleontological evolutionary sequences. One thousand replicate sequences were simulated for each combination of parameter values investigated. If the estimation procedure is unbiased, then the average of parameter estimates over these replicate sequences should be very close to the true generating values of the parameters.

Estimating μ_{step}

Estimation of μ_{step} was assessed over a range of positive and negative values (-0.1, -0.01, 0, $+0.01$, $+0.1$), with σ^2_{step} held constant at 0.1. For each value of μ_{step}, 1000 random walk sequences were generated and μ_{step} and σ^2_{step} were estimated for each on the basis of the maximum-likelihood procedure outlined above.

The estimation of μ_{step} is summarized in Figure 3, which shows histograms of parameter estimates for each of the simulated values of μ_{step}. Average parameter estimates are always very close to the true values, indicating that the procedure yields unbiased estimates of μ_{step}. Note, however, that estimates of μ_{step} span quite a large range of values, and in some instances even the inferred direction of the trend is incorrect, especially for the very subtly trended sequences ($\mu_{step} = \pm 0.01$). Nevertheless, as the quantity of data increases, this procedure can be expected to converge to the correct answer despite this variability.

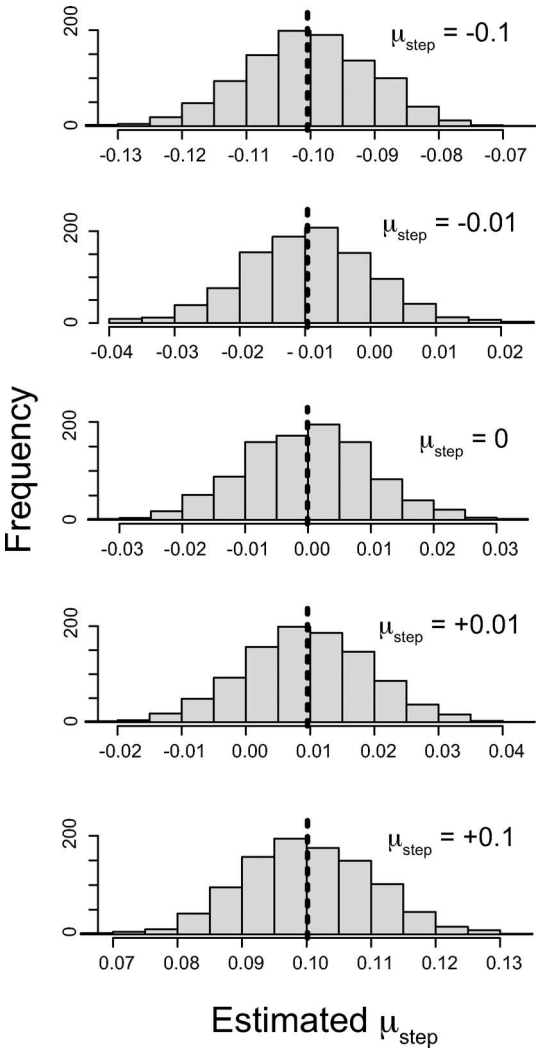

FIGURE 3. The sampling distribution of μ_{step} when estimated according to the maximum-likelihood method outlined in the text. Shown are the results for five different values of μ_{step}, corresponding to sequences (n = 20 evolutionary transitions) that are strongly directional (-0.1, $+0.1$), weakly directional (-0.01, $+0.01$) and nondirectional ($\mu_{step} = 0$). Dotted lines shows the mean of the 1000 replicates for each value of μ_{step}. In all cases the mean estimated μ_{step} is very close to the true generating value, indicating that the estimation procedure is unbiased.

Estimating σ^2_{step}

Estimation of σ^2_{step} was explored over several orders of magnitude (0.01, 0.1, 1.0, and 10 times V_p) with μ_{step} fixed at zero (the results do not differ for other values of μ_{step}). As in the previous section, 1000 replicate random walks were generated for each parameter value, and μ_{step} and σ^2_{step} were estimated for each simulated sequence.

The distribution of estimated σ^2_{step} values for each parameter value is shown in Figure 4. The average estimate of σ^2_{step} is always very close to, but slightly less than, the true parameter value. Specifically, the average estimated

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

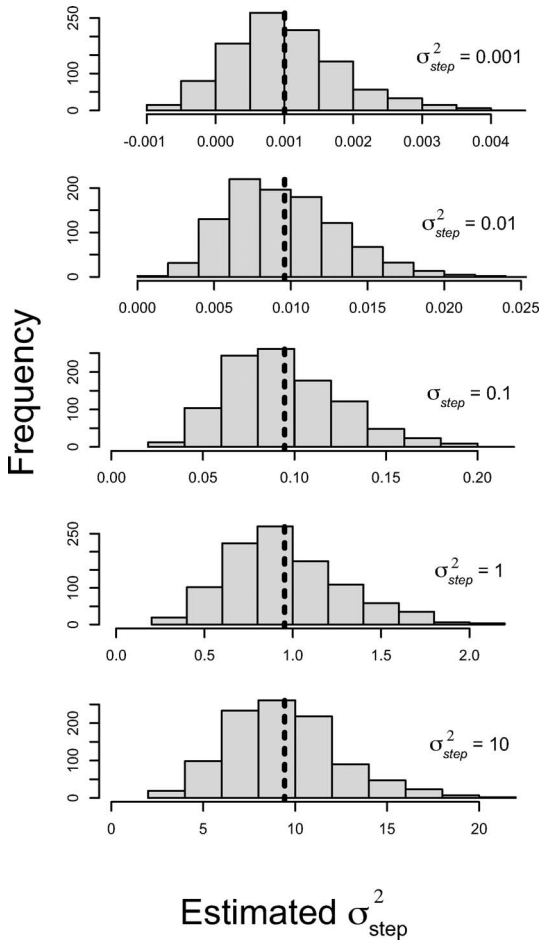

FIGURE 4. The sampling distribution of σ^2_{step} when estimated according to the maximum-likelihood method outlined in the text. Shown are the results for five different values of σ^2_{step}, increasing in magnitude from the top ($\sigma^2_{step} = 0.001$) to the bottom panel ($\sigma^2_{step} = 10$). Dotted lines shows the mean of the 1000 replicates for each value of μ_{step} (each sequence consisted of $n = 20$ evolutionary transitions). Means of the σ^2_{step} estimates tend to be very close to, but slightly less than, the true generating value, indicating a slight bias (see text for details).

σ^2_{step} in these simulations is usually about 95% of the true value. Because these sequences each included 20 evolutionary transitions (subsampled from 1000 original steps), this underestimation calls to mind the well known $(N - 1)/N$ bias of sample variances ($[20 - 1]/20 = 0.95$). In general, $(N - 1)/N$ bias occurs because deviations are slightly underestimated when computed from observed rather than true means (Bulmer 1979: p. 130). Two pieces of evidence suggest the same effect occurs when estimating the variance of the step distribution. First, over a wide range of sequence lengths, the estimation bias consistently approaches $(N - 1)/N$. Second, if we plug in to the likelihood calculations the true mean of the step distribution (rather than estimating it from the data), the $(N - 1)/N$ underestimation disappears and σ^2_{step} estimates are unbiased (unpublished results). This suggests that this small bias is a general feature of estimating variances, rather than a property of the specific method outlined here. This $(N - 1)/N$ bias in σ^2_{step} is greatest for short evolutionary sequences and it approaches zero as sequences get longer and longer. Even in short sequences, however, this bias is generally small relative to the typical sampling error of σ^2_{step}. In most practical situations, the estimated σ^2_{step} will differ from its true value more because of imprecision than because of bias.

At very low values of σ^2_{step}, two aspects of the distribution of estimated σ^2_{step} values warrant mention. First, for reasons that are not entirely clear, the $(N - 1)/N$ bias becomes weaker, and estimates of σ^2_{step} are nearly unbiased. Second, when σ^2_{step} is so low, some simulated sequences yield maximum-likelihood estimates of σ^2_{step} that are less than zero (Fig. 4). Although σ^2_{step}, like all variances, cannot truly be negative, estimates of σ^2_{step} can be less than zero if there is less difference among adjacent means than would be expected by sampling error alone. These negative estimates of σ^2_{step} pose no practical problem because they would implicitly be considered zero (as is done with negative variance estimates arising in other contexts, for example ANOVA).

Sequence Incompleteness

In any empirical paleontological evolutionary sequence, only a very small fraction of the populations existing in a lineage will ever be sampled. Roopnarine and colleagues (1999), citing a proof presented by Schroeder (1991), argued that all evolutionary sequences tend increasingly to resemble unbiased random walks as sequence completeness decreases. This is a potentially very important effect, because it would greatly limit our ability to infer directionality in paleontological sequences. It should be noted that statistical power certainly decreases as a sequence is subsampled;

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

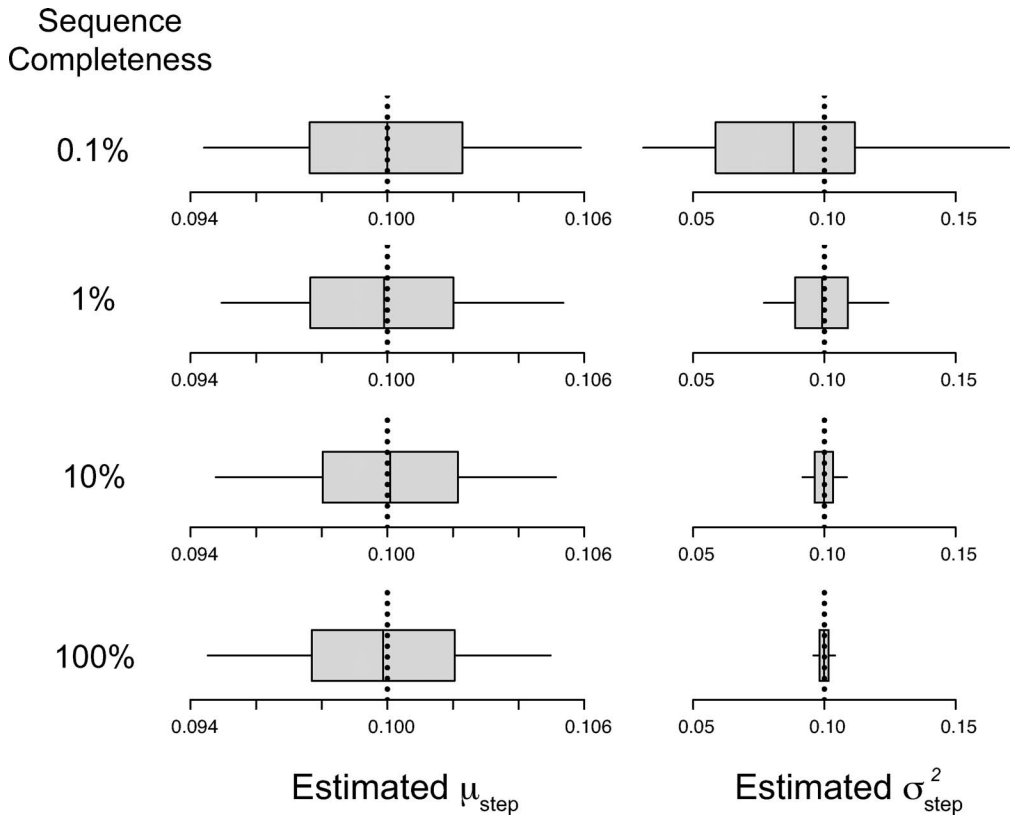

FIGURE 5. Boxplots showing sampling distribution of μ_{step} (left) and σ^2_{step} (right) when estimated from evolutionary sequences of varying levels of completeness. In all cases, the true sequence had 10,000 steps; of these, 0.1% (10 steps), 1% (100 steps), 10% (1000 steps), and 100% (10,000 steps) were sampled at random and used to estimate μ_{step} and σ^2_{step}. True values of both μ_{step} and σ^2_{step} were 0.1 for all simulations (dotted lines). Boxes indicate the middle two quartiles of the estimates, with the median indicated by a vertical bar, and the total range by the horizontal lines extending from the boxes.

with less information, rival hypotheses will be less readily discriminated. This effect is essentially universal and unavoidable. More damaging for the present approach would be if sampling intensity systematically altered parameter estimates, for example by shifting μ_{step} toward zero with increasing incompleteness. However, the results of the previous section do not seem to support this view, at least when the method described here is used. The sequences used to infer μ_{step} were incomplete, including only 20 samples from the original 1000-step random walk. Despite this incompleteness, estimates of μ_{step} were unbiased (Fig. 3). Moreover, although the proof described by Schroeder (1991) shows that the autocorrelation structure of a sequence becomes increasingly like that of an unbiased random walk with decreasing sequence completeness, it does not necessarily imply that the directionality of a sequence becomes similarly degraded. As a result, it is important to evaluate the effects of subsampling on parameter estimation empirically.

To explore the effects of sequence incompleteness on parameter estimation, I systematically varied sequence completeness and tested for its influence on the μ_{step} estimation for biased random walks ($\mu_{step} = 0.1$, $\sigma^2_{step} = 0.1$). Sequences of 10,000 steps were subsampled randomly at completenesses ranging from 0.1% to 100%. The results of these simulations (Fig. 5) show that estimates of μ_{step} are still unbiased and in particular, they do not converge on zero as completeness decreases. In fact, average μ_{step} estimates are indistinguishable from their true values at all levels of sequence completeness, including when only

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

ten of 10,000 populations (0.1%) are sampled. In addition, aside from the $(N - 1)/N$ bias, σ^2_{step} estimates are also not systematically altered by sequence incompleteness, although estimation error markedly increases with incompleteness (Fig. 5). The same pattern holds at every level of subsampling that I have simulated (at least to 0.0001%, unpublished results). Although sequence incompleteness naturally decreases precision, parameter estimates are not biased, suggesting that two key aspects of evolutionary dynamics—directionality (μ_{step}) and volatility (σ^2_{step})—can be inferred accurately in spite of pervasive incompleteness in the fossil record.

Method Performance Summary

On the whole, the simulation results suggest that this method obtains parameter estimates for the random walk model that are unbiased (μ_{step}) or nearly so (σ^2_{step}). These good estimation properties hold even when the fossil sequences include only a small percentage of the steps in the true sequence, suggesting that this approach can be applied to empirical paleontological sequences, all of which are undoubtedly incomplete.

Extending the Method

The previous sections have focused on the statistical properties of the general random walk model. Because of its simplicity and prominence in the paleontological literature, it is a natural and convenient starting point for analysis. Nevertheless, the framework advocated here is general enough to apply to a wide range of alternative models. The only requirement is that it must be possible to capture the dynamics of the model in terms of the likelihood of evolutionary transitions, as embodied by equation (6) for general random walks. Of the many possible extensions to this method, in this section I describe two: modeling stasis, and analyzing multiple evolutionary sequences jointly. Each of these extensions will be implemented in the empirical examples that follow.

A Model for Stasis

Stasis as a mode of phyletic evolution is not uniquely definable; workers have included under the banner of stasis a variety of situations in which evolutionary change is constrained in some way (Roopnarine 2001). As a result, there have been a diversity of approaches used to model stasis, including the imposition of reflecting barriers (Roopnarine 2001) and constraining forces (Hansen and Martins 1996; Martins and Hansen 1997) on random walks. Here, I use the simple model of stasis used by Sheets and Mitchell (2001b). This model imagines that there exists an optimum phenotype, with some variation permitted around this optimum but no accumulation of net morphological differences over time. More formally, trait means for all samples are normally distributed with a mean equal to the optimal phenotype (θ) and a variance of ω. The variance term (ω) captures the magnitude of evolutionary fluctuations around a fixed mean; when ω is zero, trait means are constant and lineages experience stasis in its strictest sense.

Following the procedure outlined for the general random walk, we can determine the expected evolutionary transition between ancestral and descendant populations in a lineage undergoing stasis modeled in this manner. This model differs from the random walk in that the expected mean step is not constant but instead varies as a function of the trait value of the ancestral population. Because the expected descendant morphology is always θ, evolutionary changes will tend to be negative when ancestral trait values are greater than θ and positive when ancestral values are less than θ. More precisely, the mean evolutionary transition from the ancestral population will always be $\theta - X_A$. The observed variance of these transitions will be equal to the evolutionary variance (ω) plus the sampling error associated with estimating the descendant mean trait value (V_{PD}/n_D). Substituting these terms for the mean and variance of a normal variate (eq. 4) and taking the natural logarithm yields the log-likelihood of a single transition under the stasis model:

$$\ell(\theta, \omega) = C - \frac{1}{2}\log\left(\omega + \frac{V_{PD}}{n_D}\right) - \frac{(\Delta X - [\theta - X_A])^2}{2(\omega + V_{PD}/n_D)}. \quad (9)$$

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

When sampling error does not change through a sequence, maximum-likelihood parameter estimates can be computed directly (Appendix). More generally, equation (9) can be optimized numerically over all evolutionary transitions to estimate θ and ω. Note that elapsed time does not enter the log-likelihood function because evolutionary divergence does not increase with time according to this model.

Although this formulation of stasis is rather simple, it is analytically tractable and, as will be seen below, fits at least some empirical evolutionary sequences rather well.

Evolutionary Dynamics across Multiple Sequences

Paleontologists commonly compare evolutionary patterns across different environments, regions, time intervals, and phenotypic traits. Often it is of interest to document similarities and differences with respect to these factors. In the context of the current approach, this problem can be formulated as testing whether two or more sequences share the same evolutionary dynamics. Assuming a general random walk model, this is equivalent to testing whether multiple evolutionary sequences were likely to have been generated by the same μ_{step} and σ^2_{step} values. Note that we are not testing whether specific evolutionary changes are correlated, but rather whether multiple sequences could plausibly have been generated by the same underlying dynamics.

With two observed evolutionary series, there are four possible ways in which general random walk dynamics can be similar or different across sequences: (1) both μ_{step} and σ^2_{step} are shared across sequences (homogeneous dynamics); (2) both μ_{step} and σ^2_{step} are different across sequences (heterogeneous dynamics); (3) μ_{step} is shared, but σ^2_{step} is different across sequences (heterogeneous volatility); and (4) μ_{step} is different, but σ^2_{step} is shared across sequences (heterogeneous directionality).

Each of these four models may be fit to empirical data using equation (6), or a suitable modification thereof. For the homogeneous dynamics model, equation (6) is optimized over all evolutionary transitions in both sequences to compute log-likelihoods and estimate parameters. The completely heterogeneous model is fit by optimizing equation (6) separately for each sequence; the total log-likelihood of this model is the sum of the log-likelihoods of both optimized sequences. Fitting the partially homogenous models requires modifying the fitting procedure so that μ_{step} in equation (6) is the same across sequences while multiple σ^2_{step} parameters are estimated separately for each one (or vice versa). The parameter estimates resulting from the fit of these four models provide insights into the evolutionary dynamics across multiple sequences, but they do not indicate which of the models are most consistent with observed data. This task of model selection is the subject of the next section.

Choosing Among Evolutionary Models

It has been a long-standing goal in paleontology to discriminate different modes of evolution, and random walks have figured prominently in these discussions (Raup and Crick 1981; Bookstein 1987, 1988; Gingerich 1993; Roopnarine et al. 1999; Roopnarine 2001; Sheets and Mitchell 2001b). A great advantage of likelihood-based approaches is that there are well-justified means for comparing the fit of multiple candidate models to observed data. Within the likelihood framework, determining evolutionary mode becomes a problem of model selection, with each candidate mode represented by a statistical model.

If two models fit to the same data have an equal number of parameters, the model with the higher log-likelihood is preferred. The situation is more complicated when candidate models differ in complexity because more parameters will generally improve the fit of the model to the data, even if the simpler model is actually correct. Consequently, one cannot naively compare log-likelihoods to choose among models because more complicated models have an unfair advantage.

Statisticians have devised different strategies for choosing among models of differing complexity. In one approach, the simplest of the candidate models is given the status of a null model, which is rejected in favor of a more complicated model only when the rival

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

model provides a significant improvement in fit. This procedure is familiar because of its correspondence to null hypothesis testing, but it does require that one model clearly merit null status over all other candidate models.

A second approach rates each model's fit to the data according to an information criterion, and then compares all models on an equal footing on the basis of their information content. Several information metrics have been proposed, but the most commonly used is the Akaike Information Criterion (AIC), which is equal to negative two times the log-likelihood plus twice the number of free parameters (K) in the model (Akaike 1974):

$$\mathrm{AIC} = -2\ell + 2K.$$

Although it is not derived in this way, AIC can be thought of as log-likelihood penalized by the number of parameters in the model, with more complex models suffering a correspondingly larger penalty than simpler models. When the number of observations (N) is less than about 40 times the number of parameters, a modified form of the AIC is often used to avoid overfitting (Hurvich and Tsai 1989; Anderson et al. 2000):

$$\mathrm{AIC_C} = \mathrm{AIC} + (2K[K+1])/(N-K-1).$$

Faced with a set of models and a data set, one computes the AIC or $\mathrm{AIC_C}$ for each model based on the log-likelihood of its optimal solution, the number of parameters in the model, and for $\mathrm{AIC_C}$, the number of observations (N for evolutionary sequences is the number of evolutionary transitions). The model with the lowest AIC or $\mathrm{AIC_C}$ is favored as the preferred model, with no special status granted to the simplest model.

The relative plausibility of different models can be assessed through what are called Akaike weights (Anderson et al. 2000). Their computation involves first rescaling AIC or $\mathrm{AIC_C}$ values so the minimum in the set of models is equal to zero, $\Delta_i = \mathrm{AIC}_i - \min(\mathrm{AIC})$. Then, the Akaike weight for each model is computed as

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

Akaike weights sum to one across the set of included models, and are interpreted as the approximate probabilities that each model is the best of the candidate set (Anderson et al. 2000). They are a convenient measure of the weight of evidence favoring each model, and can be used as a guide to which model or models are consistent with the observed data.

For the examples that follow, I use the information criterion approach to model selection, which has become increasingly favored in recent years (Anderson et al. 2000; Connolly and Miller 2002; Johnson and Omland 2004; Posada and Buckley 2004; Foote 2005). In the context of analyzing evolutionary sequences, one of the most compelling arguments in favor of information criterion approaches is that they obviate the need to decide which model is granted privileged null status. This decision is crucial in determining the outcome of statistical analyses (Sheets and Mitchell 2001b), but it is difficult to justify favoring any one evolutionary model a priori over other candidates.

Applying the Method

Having outlined this likelihood-based approach to analyzing evolutionary sequences, I now apply it to two sets of sequences that have been previously analyzed (and reanalyzed) in attempts to infer their evolutionary mode. The first example, based on dental traits of the Eocene primate *Cantius* (Clyde and Gingerich 1994), is used to illustrate parameter estimation and the determination of evolutionary mode. The second data set, shell shape in the planktonic foraminifera *Contusotruncana* (Kucera and Malmgren 1998), exemplifies testing for directional evolution and heterogeneity among multiple sequences.

Molar Size and Shape in *Cantius*

Previous Work.—Clyde and Gingerich (1994) traced an evolutionary sequence of *Cantius* from a single stratigraphic section in Wyoming. They recorded a series of measurements related to the size (length and width) and shape (length/width ratios, hypocone angle, and the relative position of cusps) of the upper and lower first molars. Using the LRI method (log-rate versus log-interval plots [Gingerich 1993]), these authors concluded that the size

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

and shape variables evolved according to different dynamics. They found that molar size evolved in a weakly directional manner, whereas changes in all shape variables seemed to be constrained, consistent with evolutionary stasis (Clyde and Gingerich 1994). When Sheets and Mitchell (2001b) reanalyzed these data using a variation on the LRI method, they agreed that phenotypic changes in the shape variables were constrained, but suggested that evolution in molar size traits could not be distinguished from a null model of an unbiased random walk.

Data Analysis.—Before analysis, each trait was standardized by dividing all means and standard deviations by the within-sample standard deviation pooled across all samples in the sequence (the square root of the weighted average of sample variances, with weights equal to the sample sizes minus one [Sokal and Rohlf 1995: p. 182]). This standardization does not affect the relative fit of the models, but it does allow for easier interpretation of the model parameters across traits that have different units (e.g., log mm for size traits, degrees for hypocone angle). Essentially, this converts each trait to standard deviation units, similar to the calculation of evolutionary rates in haldanes (Gingerich 1993).

For each sequence of standardized trait values, I fit the three evolutionary models: general random walk, unbiased random walk, and stasis (Table 1). These models correspond to the trichotomy of evolutionary modes that have generally been recognized: directional change, nondirectional change, and stasis, respectively (Gingerich 1993; Roopnarine 2001). In an effort to dampen some of the noise from sampling error, samples represented by a single individual were omitted, and within-sample variances were pooled over all samples unless a Bartlett's test indicated significant variance heterogeneity (Sokal and Rohlf 1995).

Evolutionary trajectories for representative size (M_1 length) and shape traits (M_1 length/width ratio) are shown in Figure 6. For convenience, sequences have been shifted so that their first sample has a mean of zero. These sequences are representative of each category: all size traits increase over time but shape

TABLE 1. Reanalysis of *Cantius* data consisting of four size-related (lengths and widths of two molars) and nine shape-related measurements (length-to-width ratios, the X and Y shape coordinates for three cusps, and the hypocone angle). Shown are the number of samples in each sequence (N), the mean number of individuals measured per sample (n), followed by the maximum-likelihood parameter estimates for the general random walk and stasis models. AIC_C values and Akaike weights are given for three models: GRW (general random walk), URW (unbiased random walk), and stasis. Akaike weights for models with more than minimal support (>0.05) are in bold. Trait sequences were transformed prior to analysis by within-sample variation, converting the parameter estimates to a common scale (see text).

Trait	Type	N	n	$\hat{\mu}_{step}$	$\hat{\sigma}^2_{step}$	$\hat{\theta}$	$\hat{\omega}$	AIC_C			Akaike weights		
								GRW	URW	Stasis	GRW	URW	Stasis
M_1 length	size	32	13.8	5.10	2.29	3.95	2.82	64.64	67.12	126.94	**0.776**	**0.224**	0.000
M_1 width	size	32	13.8	2.94	0.00	2.12	1.32	49.28	50.34	105.45	**0.630**	**0.370**	0.000
M^1 length	size	18	6.5	2.67	0.00	1.87	1.17	34.48	34.45	59.80	**0.496**	**0.504**	0.000
M^1 width	size	18	6.5	2.16	0.00	2.16	0.49	31.19	30.28	49.17	**0.388**	**0.612**	0.000
M_1 L/W	shape	32	13.8	−0.10	1.41	0.69	0.09	56.97	54.69	39.74	0.000	0.001	**0.999**
M^1 L/W	shape	18	6.5	0.26	0.86	−0.78	0.05	43.20	40.63	27.52	0.000	0.001	**0.998**
entoconid (X)	shape	20	7.7	1.12	0.00	−0.73	0.00	33.11	31.16	17.15	0.000	0.001	**0.999**
entoconid (Y)	shape	20	7.7	0.40	0.32	0.47	0.00	35.21	32.76	19.50	0.000	0.001	**0.998**
metaconid (X)	shape	20	7.7	0.05	5.06	−0.59	0.09	48.19	45.67	29.92	0.000	0.000	**1.000**
metaconid (Y)	shape	20	7.7	−0.07	1.62	0.40	0.12	46.41	43.90	31.92	0.001	0.002	**0.997**
paraconid (X)	shape	20	7.7	−1.18	0.30	−0.08	0.07	43.22	41.37	26.35	0.000	0.001	**0.999**
paraconid (Y)	shape	20	7.7	1.04	0.23	0.35	0.04	42.61	40.51	23.81	0.000	0.000	**1.000**
hypocone angle	shape	17	6.7	0.33	6.16	−0.97	0.16	49.97	47.35	34.05	0.000	0.001	**0.998**

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

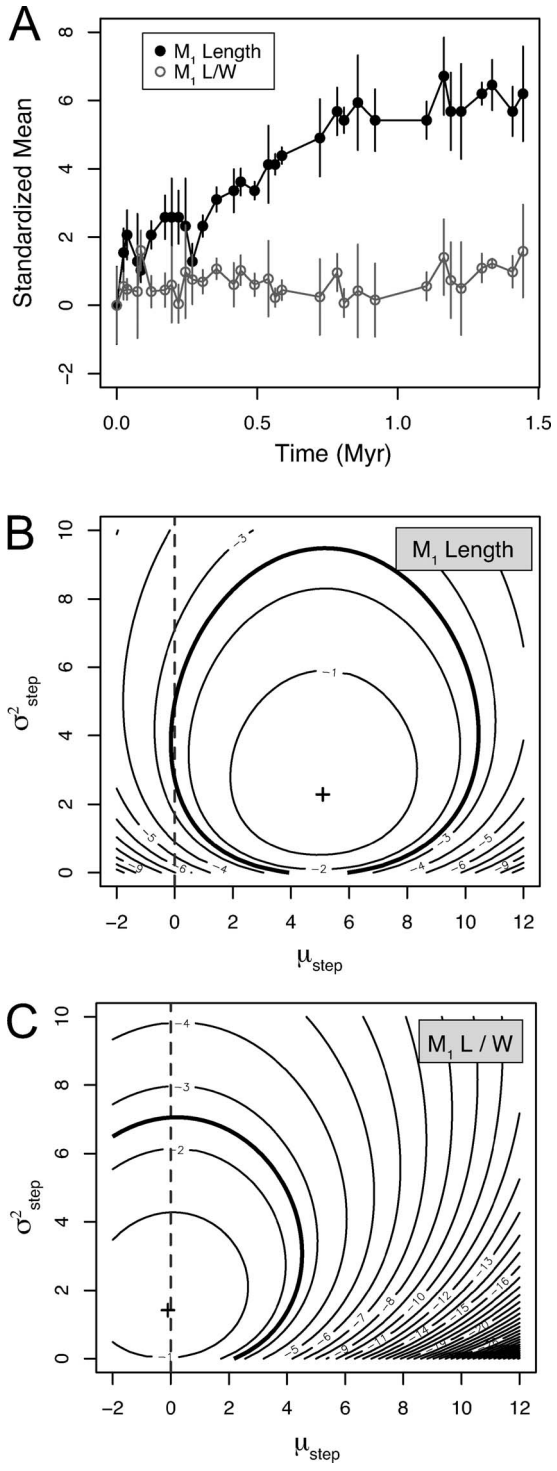

FIGURE 6. Analysis of tooth measurements in *Cantius*. A, Evolutionary sequences of a size-related trait (M_1 length) and a shape trait (M_1 L/W ratio). Dots indicate population means, with approximate 95% confidence intervals. Sequences were standardized by within-sample variance and shifted so the first sampled point has a mean of zero, as described in the text. Time scale is in Myr counting forward from the first sample. B, C, Log-likelihood surface for estimates of the parameters (μ_{step} and σ^2_{step}) of the general random walk model. B, M_1 length. C, M_1 L/W. Cross (+) indicates position of the maximum-likelihood estimate, and thin contours indicate the decrease in log-likelihood from this optimuum. The thick contour outlines the 95% joint confidence region. Solutions corresponding to an unbiased random walk are indicate by the gray dotted line at $\mu_{step} = 0$.

traits show little net change. Size increases are substantial, amounting to approximately six standard deviation units for M_1 length (Fig. 6A).

The support for various parameter values for the general random walk model can be visualized with a contour plot of the log-likelihood surface (Fig. 6B,C). The *x*- and *y*-axes of these plots correspond to estimates of μ_{step} and σ^2_{step}, and the surface height indicates the log-likelihood of each parameter combination. Each of these plots has a single peak corresponding to the maximum-likelihood estimates of μ_{step} and σ^2_{step} (indicated by a ''+''), with contour lines showing unit decreases in log-likelihood from the optimum. The thicker contour in these plots outlines solutions within 2.5 units of log-likelihood from the peak. For two-parameter models, this contour encloses the joint 95% confidence region for the parameters (see Kalinowski and Taper 2005). For M_1 length, this confidence region lies almost completely within the parameter space of positive μ_{step} values, so nearly all solutions within the 95% confidence interval imply a directional increase in this trait (Fig. 6B). The best estimate for μ_{step} is 5.10 standard deviation units per million years (SD/Myr); this is the expected net increase in M_1 length for every million years of evolution (consistent with Fig. 6A). In contrast to M_1 length, the log-likelihood surface for the length to width ratio of M_1 peaks where μ_{step} is very close to zero (−0.10 SD/Myr; Fig. 6C), indicating essentially nondirectional change. These differences in μ_{step} between size and shape variables are consistent for all traits (Table 1). All of the size traits show stronger directional change ($|\mu_{step}| > 2$) than all of the shape traits ($|\mu_{step}| < 2$), supporting the interpretation that for this lin-

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

eage segment, size has evolved in a more directional manner than shape.

Evolutionary mode can be assessed more formally by comparing the AIC_C values and Akaike weights of each of the three models (Table 1). A striking pattern emerges: all size traits show a mixture of support for the unbiased and general random walk models, whereas all shape variables overwhelmingly favor the stasis model (all Akaike weights $>$0.99; Table 1). Among size variables, dimensions of the lower molar favor a model of directional evolution (general random walk), but there is also moderate support for the unbiased random walk (Table 1). Upper molar dimensions nearly equally support unbiased and general random walk models; the lack of discrimination between these two models may partially reflect the fewer samples and smaller samples sizes for these traits (Table 1). For shape variables, the parameter estimates of the stasis models indicate evolutionary optima all very close to the starting value of zero, with only modest evolutionary variance around the optima (all less than 20% of within-sample variance; Table 1).

Overall, these results provide further quantitative support for Clyde and Gingerich's (1994) conclusions that size change was weakly directional and shape change was strongly constrained. It is also true that evolution in the size-related traits was not so strongly directional as to reject an unbiased random walk model (Sheets and Mitchell 2001b); Akaike weights for the unbiased random walk are at least moderately large for all size variables (Table 1), and so this model cannot be rejected as a candidate explanation. Under the present approach, however, the unbiased random walk is not favored by default as the null hypothesis. Instead, each model is considered equally and retained as a possible explanation to the degree that it is consistent with the data. To varying degrees specific to each of the size traits, these data are consistent with both general and unbiased random walk models.

Directionality and Geographic Heterogeneity in *Contusotruncana*

Previous Work.—Kucera and Malmgren (1998) documented evolutionary changes during the last several million years of the Cretaceous in the planktonic foraminifera lineage *Contusotruncana*. Of particular interest is the shape variable shell conicity, which showed coincident increasing trends at sites in the North Atlantic (DSDP 384) and South Atlantic (DSDP 525) Oceans (Fig. 7). When these authors applied Bookstein's (1987) scaled-maximum test, the null model of an unbiased random walk was rejected at one site (DSDP 384, $p = 0.05$) but not the other (DSDP 525, $p = 0.25$). Despite these mixed tests results, Kucera and Malmgren (1998) favored a model of directional change for this lineage, a conclusion that was later supported by a different analytical approach (Roopnarine 2001).

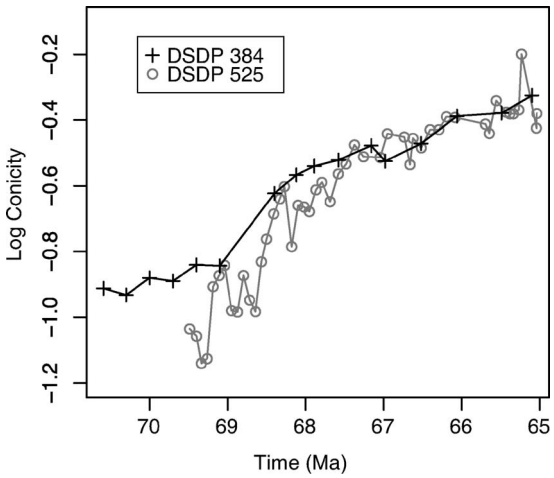

FIGURE 7. Evolutionary sequences in shell conicity for the foraminifera *Contusotrucana* at two sites: DSDP 384 (North Atlantic, black crosses), and DSDP 525 (South Atlantic, gray open circles). Note the higher volatility around the directional trend at site DSDP 525.

Data Analysis.—Conicity values were log-transformed to decrease variance heterogeneity among samples; standardization by within-sample variation is not necessary here because no comparisons are being made across traits. This reanalysis focuses on testing whether the evolutionary dynamics of shell conicity are significantly directional and the same at both sampled localities. These two issues were considered concurrently by fitting six models to these data. The six models differed in their directionality (general versus unbiased random walk), and homogeneity of process across the two localities. These mod-

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

els are listed in Table 2 and, in order of decreasing complexity, include (1) directional change and completely heterogeneous dynamics, (2) directional change and heterogeneous volatility, (3) directional change and heterogeneous directionality, (4) directional change and completely homogeneous dynamics, (5) nondirectional change and heterogeneous dynamics, and (6) nondirectional change and homogenous dynamics. The most complex model has four parameters (separate μ_{step} and σ^2_{step} for each site; model 1), whereas the simplest has just one parameter (σ^2_{step} shared across both sites, model 6). I compared the relative performance of all six models using their log-likelihoods, AIC_C values and Akaike weights.

Results.—There was great variation in the statistical support received by the six candidate models (Table 2). A directional mode of evolution is strongly supported; the two nondirectional models (5, 6) have AIC_C values much higher (= worse fit) than the best-supported directional models, and together account for a negligible proportion of the Akaike weights (Table 2). These results strongly corroborate the conclusions of previous workers that the shell conicity in this lineage evolved in a significantly directional manner (Kucera and Malmgren 1998; Roopnarine 2001).

Of the four directional models, the two that allow σ^2_{step} to differ across sites (models 1, 2) are much better supported than models that constrain σ^2_{step} to be the same (models 3, 4). This provides strong evidence that volatility of the evolutionary sequences—the magnitude of scatter around the directional trend—differed strongly at the two sites, with fluctuations being much larger at the South Atlantic site (DSDP 525). The model with the lowest AIC_C and highest Akaike weight holds μ_{step} constant across sites but allows σ^2_{step} to vary (model 2). The scenario of completely heterogeneous dynamics (model 1) also has moderate Akaike weight, but it provides almost no increase in log-likelihood relative to model 2 in exchange for its additional parameter. In fact, it is impossible for the Akaike weight of model 1 to be too much smaller than that of model 2 because of how these weights are computed. With nested models (i.e., one mod-

TABLE 2. The evolution of shell conicity in *Contusotruncana* at two sites (DSDP 384, North Atlantic Ocean; DSDP 525, South Atlantic Ocean). Each row corresponds to a model fit to the two evolutionary sequences. Models either allowed for directional evolution (GRW, general random walk), or not (URW, unbiased random walk). In addition, the models differed in terms of the homogeneity of dynamics across the two sites. In columns 3 and 4, "same" indicates that the parameter in question (μ_{step} or σ^2_{step}) was constrained to be equal at the two sites, and "diff" means that the parameter was different, i.e., estimated separately at each site. *K* indicates the number of parameters in the model, and ℓ is log-likelihood. Parameter estimates are listed for the two models that provide reasonably good fits to the observed data (as indicated by bold Akaike weights). Subscripts for parameters indicate locality names.

No.	Model type	μ_{step}	σ^2_{step}	*K*	ℓ	AIC_C	Akaike weights	Parameter estimates
1	GRW	diff	diff	4	74.43	−140.15	**0.238**	$\mu_{384} = 0.122$, $\mu_{525} = 0.117$; $\sigma^2_{384} = 0.002$, $\sigma^2_{525} = 0.066$
2	GRW	same	diff	3	74.42	−142.44	**0.745**	$\mu_{step} = 0.122$; $\sigma^2_{384} = 0.002$, $\sigma^2_{525} = 0.066$
3	GRW	diff	same	3	64.12	−121.82	0.000	—
4	GRW	same	same	2	64.12	−124.02	0.000	—
5	URW	—	diff	2	69.53	−134.85	0.017	—
6	URW	—	same	1	62.99	−123.91	0.000	—

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

el is a special case of another), the log-likelihood of the simpler model cannot exceed that of the more complex one. When the simpler model is actually correct, it will have nearly the same log-likelihood as the more complex model, and so differences in AIC are determined only by the number of parameters in the model. With only one parameter difference, the maximum AIC advantage of simpler model is 2, which corresponds to a maximum Akaike weight ≈0.73 (the exact limit depends on N as well as K for the AIC_C). This effect is unimportant here because models 1 and 2 yield almost identical parameter estimates (Table 2). Both models find that directionality is the same (or nearly so) across sites, but volatility is over 30 times higher at DSDP 525.

There is a multitude of possible explanations for the observed difference in volatility between these two localities. One possibility is that dense sampling at DSDP 525 detects fluctuations that are missed at the more coarsely sampled DSDP 384. However, when the sequence at DSDP 525 is subsampled at a resolution similar to DSDP 384, the same qualitative differences remain (results not shown). Another possibility is that environmental fluctuations were larger in amplitude at DSDP 525. However, these sites were at similar paleolatitude, and there does not appear to be an obvious oceanographic reason why the South Atlantic would be more variable than the North Atlantic Ocean during this interval (B. Huber personal communication 2006). Michal Kucera (personal communication 2006) has suggested two additional possibilities. First, sedimentation rate was slower at DSDP 384 and consequently its samples are more time-averaged, dampening evolutionary fluctuations at that site. Second, the location of DSDP 525 near the periphery of *Contusotruncana*'s geographic range suggests that environmental fluctuations may result in less optimal conditions for these foraminifera, which could translate into greater phenotypic variation. Regardless of the underlying cause or causes, the analyses presented here demonstrate that there is an interesting geographic pattern to be explained.

Discussion

The approach described in this paper, at minimum, provides a means to fit statistical models to paleontological sequences. More broadly, it can be a starting point for developing a statistical framework with which to evaluate hypotheses about evolutionary mode. Because this method is implemented via maximum likelihood, it inherits all the benefits of this approach, including estimators with good statistical properties (Le 1992) and rigorous criteria for choosing among competing models. Simulation experiments demonstrate that this procedure provides reliable estimates of the random walk parameters, even when evolutionary sequences are very incomplete.

The likelihood functions developed here explicitly take into account sampling error in estimating trait means in fossil populations. Although it may seem a mundane concern, sampling error can account for a surprisingly large proportion of the observed differences between populations in an evolving sequence (see also Kinnison and Hendry 2001). Ignoring this source of variation will inflate estimates of evolutionary parameters by conflating fluctuations due to sampling with true evolutionary differences. For example, ignoring sampling error in the *Cantius* analysis yields a σ^2_{step} estimate for M_1 length that overestimates the maximum likelihood estimate by over 500%. In addition to its substantial magnitude, sampling error is not neutral with respect to determination of evolutionary mode. Most of these tests work by comparing divergence over short time scales (e.g., between adjacent samples) to those over longer time scales (e.g., the maximum divergence over a whole or large part of a sequence); this is true for the scaled maximum test (Bookstein 1987), the LRI method (Gingerich 1993), and methods based on computation of Hurst exponents (Roopnarine et al. 1999). Sampling error often substantially inflates short-term divergence, but it has a much smaller effect on the total or maximum divergence in a sequence because it does not accumulate over time. This has the effect of making evolutionary sequences appear less directional, and more like stasis, than they actually are. This

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

factor, in conjunction with the low power of these tests (Roopnarine et al. 1999; Sheets and Mitchell 2001b), may be part of the reason why tests rarely reject the null hypothesis of an unbiased random walk in favor of directional change. Moreover, given that rounding error can influence the outcome of random walk tests (Sheets and Mitchell 2001b), sampling error—which is similar in effect but always much larger—must be even more important.

Within this statistical framework, I have focused primarily on the specific model of a general random walk. This model is probably the simplest that allows the magnitude and direction of evolution to vary in a realistic manner. Reducing the complex network of causes that influence trait evolution to one or two generating parameters is surely a drastic simplification of reality. Random walks, like all models, are certainly not true in the sense of providing a complete account of the phenomenon under study (Anderson et al. 2000). Nevertheless, simplified models can be very useful for extracting interesting aspects of complex systems. For example, regardless of the complexities of the underlying evolutionary mechanisms, μ_{step} is likely to capture something important about the directionality of trait evolution. Another rationale for starting with random walks is that they are already the basis for analyzing phenotype in a phylogenetic context. Nearly all commonly used comparative methods are based on the unbiased random walk model (Felsenstein 1985; Martins 1999; Garland and Ives 2000). The same parameters, μ_{step} and σ^2_{step}, can be estimated from morphological differences among terminal taxa in a phylogeny (Martins 1994; Pagel 2002), although μ_{step} cannot be estimated if all tips are of the same age, as in studies of exclusively extant taxa. Random walk models therefore provide a common framework to evaluate evolutionary patterns in fossil sequences and in phylogenetically related modern species.

For all their reasonableness, random walk models may at times fail to adequately capture the dynamics of an evolving sequence or sequences. A great advantage of this approach is that one can then specify one or more alternative models of evolutionary change and compare their fit to the data in a rigorous way. In this paper, I have considered the three modes of evolutionary change commonly considered in paleontological studies: directional change (general random walk), nondirectional change (unbiased random walk), and stasis. Each of these three models can be fit to paleontological sequences using maximum likelihood, and the relative support of each can be assessed using well-established statistical means such as AIC_C scores and Akaike weights.

This approach shares with two previously proposed methods (Gingerich 1993; Roopnarine 2001) the strategy of assessing the merits of different modes of evolution without granting one of them privileged null status. The LRI method (Gingerich 1993) plots rates versus interval length on a double log scale and infers evolutionary mode on the basis of the resulting slope. Slopes of zero indicate completely directional change, stasis produces slopes near negative one, and (unbiased) random walks are intermediate (Gingerich 1993). This method has been criticized (Sheets and Mitchell 2001a; Roopnarine 2003), but at least for the *Cantius* data, it yields conclusions that are consistent with the likelihood approach adopted here. Roopnarine's (2001) method is based on Hurst exponents (which can be related to the slope of LRI plots) computed using a moving window through an evolutionary sequence. Application of this method to the *Contusotruncana* shell conicity data finds evidence for directional evolution (Roopnarine 2001), a finding that is also supported in the present reanalysis.

It is reassuring that independent and reasonable methods produce consistent results when applied to the same data. Although a comparative analysis of the performance of these different approaches is clearly beyond the scope of this paper, the current methodology does have several advantages. The parameter estimates that result from model fits are relevant to biological hypotheses (e.g., directionality), confidence intervals can be computed readily (Kalinowski and Taper 2005), and the quantitative support candidate evo-

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

lutionary models can be assessed through convenient and well-established means.

Another key feature of the current approach is that it is extendable. Many other kinds of models can be tested, corresponding to a wide range of hypotheses about patterns and causes of phyletic evolution. In the present study, the approach was extended to consider the homogeneity of evolutionary dynamics across multiple evolutionary sequences. Although the differences between sequences in this case was geographic, the same logic can be applied to test whether evolutionary patterns differ across environments, lineages, traits, and modes of life (e.g., benthic versus pelagic [McKinney 1985]).

The current approach could also be modified for situations in which the direction of evolutionary change is influenced by an extrinsic variable. For example, it has been argued that body size evolution may be shaped by environmental factors such as temperature (Davis 1981; Smith et al. 1995; Hunt and Roy 2006), oxygen concentration (Chapelle and Peck 1999; Kaiho 1999; McClain and Rex 2001), and ocean structure (Schmidt et al. 2004; Finkel et al. 2005). Hypotheses such as these can be evaluated with a likelihood function in which mean and variance of the step distribution are not constant, but instead vary as functions of environmental parameters. The functional form of the relationship can be specified on the basis of prior knowledge, or a range of functional forms can be fit as separate models and the usual model selection methods used to determine the most promising candidate.

Alternatively, the step distribution could be a function of the ancestral morphology, as in structured state space models (Alroy 2000). This family of models allows for different evolutionary rules to dominate in different regions of the morphological space. The stasis model presented earlier is a simple example of this kind of model; its dynamics are not the same for all phenotypes but instead are determined by where the ancestral morphology is relative to the optimum.

Finally, yet another extension to this method would allow for evolutionary dynamics to change within a sequence. This has been a very common scenario to consider, starting with the suggestion of ''punctuated gradualism'' in a planktonic foraminifera lineage (Malmgren et al. 1983), and continuing through the several reanalyses of this same data set (Bookstein 1987; MacLeod 1991; Roopnarine 2001). This view is also implicit in earlier debates over whether to interpret changes within lineages as gradual or as a series of punctuation events (Gould and Eldredge 1977). One way to model this notion is to imagine that a sequence begins with a set of generating parameters, which then shift to a second set of parameters at some point in the evolutionary sequence, with further shifts possible. If there is a prior hypothesis as to where in the sequence this shift should occur (e.g., at a boundary, or an environmental change), then the model-selection procedures apply as usual. However, if the shift point is chosen on the basis of the pattern of phenotypic changes in the sequence, these methods must be modified because they are designed to evaluate models erected independently of the data. When one model is derived from observed data, model-selection routines will favor that model more than they should (this is akin to separating tall people from short and then testing if their growth rates differ). One solution to this problem could be to estimate the shift point as a free parameter from the data, rather than from visual inspection of the sequence of trait values (see Roopnarine 2001 for a different approach to this same problem).

In this paper, I have outlined an approach that represents evolutionary hypotheses as statistical models. Historically, assessing evolutionary mode in phyletic sequences has been one of the paleontological endeavors most in need of an explicit, model-based approach. Early in the debate over gradualism versus punctuated equilibria, there was much disagreement about whether specific evolutionary series showed gradational or pulsed patterns of change (Gould and Eldredge 1977). An approach based on likelihood analysis of specified models offers a way out of the subjectivity surrounding these and other debates by providing the common scale upon which scientific hypotheses about evolutionary mode can be weighed.

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

Summary

1. In a general form of the random walk model, evolutionary changes are drawn randomly at each time interval from a distribution of evolutionary transitions, or ''steps.'' The long-term dynamics of this model are influenced solely by the mean (μ_{step}) and variance (σ^2_{step}) of the distribution of evolutionary steps; the former determines the directionality of evolutionary sequences, and the latter governs their volatility.
2. Simulations show that two parameters of the general random walk (μ_{step} and σ^2_{step}) can be estimated reliably from paleontological sequences by using maximum likelihood. Moreover, these parameter estimates are not systematically altered by completeness with which evolutionary sequences are sampled.
3. Two extensions to the model are developed: modeling evolutionary stasis and analyzing evolutionary dynamics across multiple sequences. Stasis, random walks, and a wide range of other evolutionary models can be evaluated under this approach to determine which are most consistent with observed data. Model selection can be accomplished through well-established means, including use of information criteria (such as AIC) that compare all candidate models on an equal footing.
4. Reanalysis of Clyde and Gingerich's *Cantius* data supports their original interpretation that molar size evolution was weakly directional, whereas molar shape evolution was constrained (stasis).
5. Shell conicity in *Contusotruncana* evolved in a significantly directional manner at two sites. In addition, evolutionary volatility (as measured by σ^2_{step}) was much higher at one of the two sites, indicating heterogeneity of evolutionary process within this species' range.

Acknowledgments

I thank M. Foote for useful comments on an earlier version of the manuscript, and for prompting me to derive analytical maximum-likelihood estimators. S. Wing, P. Roopnarine, and two anonymous reviewers helped steer the manuscript toward greater accessibility; one reviewer in addition offered the particularly useful suggestion of using Akaike weights to summarize model support.

Literature Cited

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723.

Alroy, J. 2000. Understanding the dynamics of trends within evolving lineages. Paleobiology 26:319–329.

Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. Null hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management 64:912–923.

Bookstein, F. L. 1987. Random walk and the existence of evolutionary rates. Paleobiology 13:446–464.

———. 1988. Random walk and the biometrics of morphological characters. Evolutionary Biology 9:369–398.

Bulmer, M. G. 1979. Principles of statistics. Dover, New York.

Chapelle, G., and L. S. Peck. 1999. Polar gigantism dictated by oxygen availability. Nature 399:114–115.

Clyde, W. C., and P. D. Gingerich. 1994. Rates of evolution in the dentition of early Eocene *Cantius*: comparison of size and shape. Paleobiology 20:506–522.

Connolly, S. R., and A. I. Miller. 2002. Global Ordovician faunal transitions in the marine benthos: ultimate causes. Paleobiology 28:26–40.

Cunningham, C. W., K. E. Omland, and T. H. Oakley. 1998. Reconstructing ancestral character states: a critical reappraisal. Trends in Ecology and Evolution 13:361–366.

Davis, S. J. M. 1981. The effects of temperature change and domestication on the body size of Late Pleistocene to Holocene mammals of Israel. Paleobiology 7:101–114.

Edwards, A. W. F. 1992. Likelihood. Johns Hopkins University Press, Baltimore.

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist 125:1–15.

Finkel, Z. V., M. E. Katz, J. D. Wright, O. M. E. Schofield, and P. G. Falkowski. 2005. Climatically driven macroevolutionary patterns in the size of marine diatoms over the Cenozoic. Proceedings of the National Academy of Sciences USA 102:8927–8932.

Foote, M. 2003. Origination and extinction through the Phanerozoic: a new approach. Journal of Geology 111:125–148.

———. 2005. Pulsed origination and extinction in the marine realm. Paleobiology 31:6–20.

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: 346–364.

Gingerich, P. D. 1976. Paleontology and phylogeny: patterns of evolution at the species level. American Journal of Science 276: 1–28.

———. 1993. Quantification and comparison of evolutionary rates. American Journal of Science 293-A:453–478.

Gould, S. J. 2002. The structure of evolutionary theory. Belknap Press of Harvard University Press, Cambridge.

Gould, S. J., and N. Eldredge. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology 3: 115–151.

Hallam, A. 1968. Morphology, palaeoecology and evolution of the genus Gryphaea in the British Lias. Philosophical Transactions of the Royal Society of London B 254:91–128.

Hansen, T. F., and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

correlation structure of interspecific data. Evolution 50:1404–1417.
Hunt, G. 2004. Phenotypic variation in fossil samples: modeling the consequences of time-averaging. Paleobiology 30:426–443.
Hunt, G., and K. Roy. 2006. Climate change, body size evolution, and Cope's Rule in deep-sea ostracodes. Proceedings of the National Academy of Sciences USA 103:1347–1352.
Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika 76:297–307.
Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19:101–108.
Kaiho, K. 1999. Evolution in the test size of deep-sea benthic foraminifera during the past 120 m.y. Marine Micropaleontology 37:53–65.
Kalinowski, S. T., and M. L. Taper. 2005. Likelihood-based confidence intervals of relative fitness for a common experimental design. Canadian Journal of Fisheries and Aquatic Science 62:693–699.
Kinnison, M. T., and A. P. Hendry. 2001. The pace of modern life II: from rates of contemporary microevolution to pattern and process. Genetica 112-113:145–164.
Kucera, M., and B. A. Malmgren. 1998. Differences between evolution of mean form and evolution of new morphotypes: an example from Late Cretaceous planktonic foraminifera. Paleobiology 24:49–63.
Le, C. T. 1992. Fundamentals of biostatistical inference. Marcel Dekker, New York.
Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. American Naturalist 136:727–741.
Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, Mass.
MacLeod, N. 1991. Punctuated anagenesis and the importance of stratigraphy to paleobiology. Paleobiology 17:167–188.
Malmgren, B. A., W. A. Berggren, and G. P. Lohmann. 1983. Evidence for puncuated gradualism in the Late Neogene *Globorotalia tumida* lineage of planktonic foraminifera. Paleobiology 9:377–389.
Martins, E. P. 1994. Estimating the rate of phenotypic evolution from comparative data. American Naturalist 144:193–209.
———. 1999. Estimation of ancestral states of continuous characters: a computer simulation study. Systematic Biology 48: 642–650.
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 149:646–667.
McClain, C. R., and M. A. Rex. 2001. The relationship between dissolved oxygen concentration and maximum size in deep-sea turrid gastropods: an application of quantile regression. Marine Biology 139:681–685.
McKinney, M. 1985. Distinguishing patterns of evolution from patterns of depostion. Journal of Paleontology 59:561–567.
Pagel, M. 2002. Modelling the evolution of continuously varying characters on phylogenetic trees: the case of Hominid cranial capacity. Pp. 269–286 *in* N. MacLeod and P. L. Forey, eds. Morphology, shape and phylogeny. Taylor and Francis, London.
Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53:793–808.
R Development Core Team. 2005. R: A language and environment for statistical computing, Version 2. 1.1. R Foundation for Statistical Computing, Vienna.
Raup, D. M. 1977. Stochastic models in evolutionary paleobiology. Pp. 59–78 *in* A. Hallam, ed. Patterns of evolution as illustrated by the fossil record. Elsevier, Amsterdam.
Raup, D. M., and R. E. Crick. 1981. Evolution of single characters in the Jurassic ammonite *Kosmoceras*. Paleobiology 7:200–215.
Raup, D. M., and S. J. Gould. 1974. Stochastic simulation and evolution of morphology—towards a nomothetic paleontology. Systematic Zoology 23:305–322.
Roopnarine, P. D. 2001. The description and classification of evolutionary mode: a computational approach. Paleobiology 27: 446–465.
———. 2003. Analysis of rates of morphologic evolution. Annual Review of Ecology and Systematics 34:605–632.
Roopnarine, P. D., G. Byars, and P. Fitzgerald. 1999. Anagenetic evolution, stratophenetic patterns, and random walk models. Paleobiology 25:41–57.
Schmidt, D. N., H. R. Thierstein, J. Bollman, and R. Schiebel. 2004. Abiotic forcing of plankton evolution in the Cenozoic. Science 303:207–210.
Schroeder, M. 1991. Fractals, chaos, power laws. W. H. Freeman, New York.
Sheets, H. D., and C. E. Mitchell. 2001a. Uncorrelated change produces the apparent dependence of evolutionary rate on interval. Paleobiology 27:429–445.
———. 2001b. Why the null matters: statistical tests, random walks and evolution. Genetica 112– 113:105–125.
Smith, F., J. L. Betancourt, and J. H. Brown. 1995. Evolution of body size in the woodrat over the past 25,000 years of climate change. Science 270:2012–2014.
Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research, 3d ed. W. H. Freeman, New York.
Trueman, A. E. 1922. The use of *Gryphaea* in the correlation of the lower Lias. Geological Magazine 59:256–268.

Appendix

Parameter Estimates for the General Random Walk Model

For the special case in which samples are evenly spaced in time and sampling error for trait means is the same for all samples, simple equations can be derived for the maximum-likelihood estimates for the parameters of the general random walk, μ_{step} and σ^2_{step}.

In order to simplify notation, let x denote an evolutionary transition in trait X ($x = X_D - X_A$). If all samples have the same phenotypic variance (V_p) and sample size (n), the sampling variance for each transition will be equal to $\varepsilon = 2V_p/n$. Making these substitutions into equation (6) yields the log-likelihood of a single evolutionary transition, x

$$\ell(\mu_{step}, \sigma^2_{step}; x) = C - \frac{1}{2}\log(t\sigma^2_{step} + \varepsilon) - \frac{(x - t\mu_{step})^2}{2(t\sigma^2_{step} + \varepsilon)}. \quad \text{(A.1)}$$

Summing equation (A.1) over a vector ($\mathbf{x}$) of N evolutionary transitions yields

$$\ell(\mu_{step}, \sigma^2_{step}; \mathbf{x}) = \sum_{i=1}^{N}\left[C - \frac{1}{2}\log(t\sigma^2_{step} + \varepsilon) - \frac{(x_i - t\mu_{step})^2}{2(t\sigma^2_{step} + \varepsilon)}\right].$$

Taking the constant and the next term outside the summation,

$$\ell(\mu_{step}, \sigma^2_{step}; \mathbf{x}) = C - \frac{N}{2}\log(t\sigma^2_{step} + \varepsilon) - \frac{1}{2(t\sigma^2_{step} + \varepsilon)} \cdot \sum_{i=1}^{N}(x_i - t\mu_{step})^2$$

then expanding the remaining summation and substituting the

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo

sum of all x_i as the mean x times N yields

$$\ell(\mu_{step}, \sigma^2_{step}; \mathbf{x}) = C - \frac{N}{2}\log(t\sigma^2_{step} + \varepsilon) - \frac{1}{2(t\sigma^2_{step} + \varepsilon)} \cdot \left(\sum_i x_i^2 - 2N\bar{x}t\mu_{step} + Nt^2\mu^2_{step}\right). \quad (A.2)$$

Parameter estimates for μ_{step} and σ^2_{step} are obtained by setting the partial derivatives of equation (A.2) equal to zero and solving. The resulting parameter estimates are

$$\hat{\mu}_{step} = \bar{x}/t, \quad (A.3)$$

$$\hat{\sigma}^2_{step} = \frac{1}{t}\left(\frac{1}{N}\sum_i x_i^2 - \bar{x}^2 - \varepsilon\right). \quad (A.4)$$

Within the parentheses, the first two terms constitute the maximum-likelihood estimate of the variance of x. As a result, equation (A.4) can also be written as

$$\hat{\sigma}^2_{step} = \frac{1}{t}(\mathrm{Var}[x] - \varepsilon). \quad (A.5)$$

Thus, the sampling error associated with estimating trait means is subtracted from the observed variance in evolutionary steps in order to estimate σ^2_{step}. Note that maximum likelihood estimate of variance from equation (A.4) is downwardly biased, accounting for the $(N - 1)/N$ bias observed in the estimating σ^2_{step} (Lynch and Walsh 1998, p. 810).

Parameter Estimates for the Stasis Model

Assuming constant sampling error, we can derive simple equations for maximum likelihood estimators of the parameters of the stasis model: θ, the trait optimum, and ω, the variance around this optimum. Again, let X denote a trait value, and x refer to an evolutionary transition in this trait ($x = X_D - X_A$). As explained in the text, the expected evolutionary transition (x) is a function of the ancestral trait value, such that the mean step is $\theta - X_A$. The step variance is equal to $\omega + \varepsilon_X$, where ε_X is the sampling variance of the descendant population (the sampling error of X_A does not contribute because we are conditioning on the observed ancestral trait value). Assuming that population means are normally distributed around θ, the log-likelihood of x is obtained by substituting the appropriate mean and variance into equation (4):

$$\ell(\theta, \omega; x) = C - \frac{1}{2}\log(\omega + \varepsilon_X) - \frac{(x - [\theta - X_A])^2}{2(\omega + \varepsilon_X)}. \quad (A.6)$$

Because $x + X_A$ is equal to X_D, the numerator of the right-most term becomes $(X_D - \theta)^2$. From this point, the steps closely parallel the preceeding section and are not shown. Briefly, equation (A.6) is summed over all evolutionary transitions, and the partial derivatives of the resulting log-likelihood function are set to zero and solved. The resulting estimators are

$$\hat{\theta} = \bar{X}_D, \quad (A.7)$$

$$\hat{\omega} = \mathrm{Var}[X_D] - \varepsilon_X. \quad (A.8)$$

These estimators are quite sensible: the optimum is estimated simply as the mean of all descendant trait values, and evolutionary variance around this optimum is estimated as the variance of descendant trait values with the contribution of sampling error removed.

Downloaded From: https://bioone.org/journals/Paleobiology on 21 Oct 2020
Terms of Use: https://bioone.org/terms-of-use Access provided by University of Oslo