

See discussions, stats, and author profiles for this publication at:  
<http://www.researchgate.net/publication/236883885>

# Measuring rates of phenotypic evolution and the inseparability of tempo and mode

ARTICLE *in* PALEOBIOLOGY · JUNE 2012

Impact Factor: 2.66 · DOI: 10.2307/41684606

---

CITATIONS

27

---

READS

86

## 1 AUTHOR:



[Gene Hunt](#)

Smithsonian Institution

62 PUBLICATIONS 1,155 CITATIONS

SEE PROFILE

# *Paleobiology*

**Measuring rates of phenotypic evolution and the inseparability of tempo and mode**

*Gene Hunt*



## Measuring rates of phenotypic evolution and the inseparability of tempo and mode

Gene Hunt

**Abstract.**—Rates of phenotypic evolution are central to many issues in paleontology, but traditional rate metrics such as darwins or haldanes are seldom used because of their strong dependence on interval length. In this paper, I argue that rates are usefully thought of as model parameters that relate magnitudes of evolutionary divergence to elapsed time. Starting with models of directional evolution, random walks, and stasis, I derive for each a reasonable rate metric. These metrics can be linked to existing approaches in evolutionary biology, and simulations show that they can be estimated accurately at any temporal resolution via maximum likelihood, but only when that metric's underlying model is true.

The estimation of generational rates of a random walk under realistic paleontological conditions is compared with simulations to that of a prominent alternative approach, Gingerich's LRI (log-rate, log-interval) method. Generational rates are estimated poorly by LRI; they often reflect sampling error more than the actual pace of change. Further simulations show that under some realistic conditions, it is simply not possible to infer generational rates from coarsely sampled populations.

These modeling results indicate a complex dependence between evolutionary mode and the measurement of evolutionary rates, and that there is unlikely to be a rate metric that works well for all traits and time scales. Compilations of paleontological and phylogenetic data indicate that all of the three rate metrics derived here show some relationship with interval length. Although there is no perfect rate metric, at present the most practical choices derive from the parameters of the stasis and random walk models. The latter, called the step variance, is particularly promising as a rate metric in paleontology and comparative biology.

Gene Hunt. Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560. E-mail: [hunte@si.edu](mailto:hunte@si.edu)

Accepted: 27 December 2011

Supplementary materials deposited at Dryad: doi: 10.5061/dryad.c1m60s84

### Introduction

Many key questions in evolutionary paleobiology hinge upon the pace at which morphology changes. Why was morphological evolution so fast during the Cambrian Explosion of animal body plans (Marshall 2006; Erwin et al. 2011)? How much faster are punctuations compared to stasis (Gould and Eldredge 1977; Gingerich 1985)? Do some aspects of morphology evolve faster than others, and do some conditions preferentially spur rapid morphological changes (e.g., Jablonski and Bottjer 1990)? Although these issues involve comparing rates of phenotypic change, only seldom do paleontologists address them by actually computing rates of morphological evolution.

Haldane (1949) formalized what is now the traditional measurement of the rate at which a quantitative trait changes:  $\Delta X/t$ , where  $\Delta X$  is the change in a trait from ancestor to descendant and  $t$  is the elapsed time between

the two. Morphological difference is typically measured in natural logarithm or standard deviation units, and elapsed time in millions of years or generations. Certain combinations of units are given names: darwins record changes in log units per Myr whereas haldanes measure change in standard deviation units per generation (Haldane 1949; Gingerich 1993).

Although simple to compute, traditional rates are seldom used because they are strongly resolution-dependent: the longer the window of observation, the slower that evolution appears to be (Gingerich 1983). Many factors have been said to contribute to this correlation, including the effects of evolutionary reversals and stasis, constraints, difficulties in detecting very slow rates, the biological implausibility of sustained and rapid changes, and the artifact of plotting a ratio against its denominator (Gingerich 1983; Gould 1984; Foote 1991; Hendry and Kinnison

1999; Sheets and Mitchell 2001a). Regardless of cause, this strong temporal scaling poses an important practical problem because rates measured over different interval lengths are not easily compared.

Computing a rate is essentially an effort to standardize magnitudes of change in a way that accounts for elapsed time. This accounting is complicated, however, because evolutionary change accumulates with elapsed time differently under different evolutionary scenarios (Gingerich 1993; Roopnarine 2003; Estes and Arnold 2007). Thus any accounting requires that a rate metric assume something about the true nature of change. Traditional rates parallel the definition of velocity in physics, distance divided by time, and in doing so they invoke a simple and directional notion of evolutionary change. But because traits do not usually evolve at constant rates over paleontological time scales, this analogy may be particularly unhelpful.

The view taken here is that rates are usefully construed as parameters in models that predict the magnitude of change over specified intervals of time. Rates are not data but rather follow from data when combined with a model of how phenotypes evolve (Bookstein 1987; Sheets and Mitchell 2001a). This model choice turns out to be an important but overlooked aspect of measuring rates of evolution, and in this paper I explore how evolutionary mode affects the measurement of evolutionary tempo. The general strategy is as follows. I start with standard models of phenotypic evolution (random walks, directional change, stasis) and derive a sensible way to measure rate within the context of each model. I then use simulations to explore how well these rates can be estimated in a variety of scenarios and compare this performance with that of a prominent alternative for measuring rates, Gingerich's LRI (log-rate, log-interval) approach. Finally, I use compilations of paleontological and phylogenetic data sets to assess the behavior of different rate metrics in real lineages and clades.

The sections of this paper are semi-independent, and in places, necessarily technical. But together they have important practical

implications for how paleontologists measure rates of phenotypic change. They suggest that it is possible to measure rates well at any temporal resolution when evolution works according to known and simple evolutionary models, but that no rate metric will be ideal when the true mode of evolution varies across trait or lineages, or when evolution deviates from simple uniform models. Moreover, under some realistic conditions, simulations show that it is simply not possible to take morphological divergence observed at typical paleontological resolutions and infer the magnitude of changes occurring from one generation to the next. Perhaps most importantly, the two most familiar tools for estimating rates—the LRI approach and traditional rate metrics such as haldanes—are both problematic, and theoretical and empirical considerations suggest that existing alternatives, including a rate parameter derived from the random walk model, are much more promising.

### Measuring Tempo for Different Modes of Evolution

#### Divergence under Different Modes of Evolutionary Change

[Gingerich \(1993; see also Roopnarine 2003\)](#) showed that the temporal scaling of traditional rate metrics differs under different evolutionary scenarios—they have different slopes on a plot of log-rate versus log-interval. Instead of looking at one rate metric across different models, here I take the reverse approach: Starting with a set of different evolutionary models, I derive for each a sensible rate metric.

To derive rate metrics we need to formulate a measure of expected change in a trait,  $X$ , as a function of time. Most models of evolution are stochastic, not deterministic, in that they predict a probability distribution of outcomes rather than a single value. The three models considered here—directional evolution, random walk, and stasis—predict that the expected changes between ancestors and descendants will be normally distributed with mean and variance that depend on the model parameters and elapsed time (Hunt 2006,

TABLE 1. Evolutionary divergence and rate metrics for three models of evolution. For each, evolutionary changes from ancestor to descendant ( $\Delta X$ ) over  $t$  generations will be normally distributed with means and variance given in the first two columns. The fourth column shows the expression for expected squared divergence, ESD, a useful measure of expected evolutionary change. The ESD for the stasis model integrates over the distribution of ancestral states ( $X_A$ ) as described in the text. The fifth column gives useful a rate metric for each model, and the last column gives the interpretation of these rate metrics.

Evolutionary model	Mean $\Delta X$	Variance $\Delta X$	Expected squared divergence, ESD	Useful rate metric	Related approach
Directional	$\mu_s t$	$\sigma_s^2 t$	$\mu_s^2 t^2 + \sigma_s^2 t$	$\mu_s^2$ *	Traditional rates
Random walk	0	$\sigma_s^2 t$	$\sigma_s^2 t$	$\sigma_s^2$	Step variance/Brownian motion rate
Stasis	$\theta - X_A$	$\omega$	$2\omega$	$\omega$	Disparity

\* Only if evolution is strongly directional.

2008a) (Table 1). Directional evolution, as modeled here, draws an evolutionary change for each generation from a distribution of evolutionary “steps.” This distribution has a mean step,  $\mu_s$ , representing the per-generation trend, and a step variance,  $\sigma_s^2$ , capturing the deviations from this trend. After  $t$  generations, changes from ancestor to descendant,  $\Delta X$ , will be normally distributed with a mean  $\mu_s t$  and variance  $\sigma_s^2 t$  (Hunt 2006). A random walk is a special case of the directional model in which  $\mu_s = 0$ , and so  $\Delta X$  according to this model will be normally distributed with

variance also equal to  $\sigma_s^2 t$  but with a mean of zero (Table 1, Fig. 1).

For stasis, I follow Sheets and Mitchell (2001b) and consider it as uncorrelated, normally distributed variation with variance  $\omega$  around a stable long-term mean,  $\theta$ . From any ancestral value,  $X_A$ , traits are expected to converge immediately to the optimum,  $\theta$ , and thus the expected change is  $\theta - X_A$ . Variance around this expectation is  $\omega$  (Table 1). Under the strictest notion of stasis,  $\omega = 0$  and the trait is truly static over time.

### A Quantitative Measure of Divergence

The three canonical models of evolution all produce normally distributed evolutionary changes. The next step is to distill from this probability distribution a measure of the magnitude of evolutionary divergence. For rates, we are usually uninterested in the direction of change, and the magnitude of difference is often expressed as the absolute value of morphological change,  $|\Delta X|$  (e.g., Hendry and Kinnison 1999; Estes and Arnold 2007; Gingerich 2009). Because absolute values are inconvenient for analysis, I will instead use the squared difference in morphology,  $\Delta X^2$ . A useful summary of these is simply the mean, or expected squared divergence (ESD, or  $E[\Delta X^2]$ ). ESD is similar to variance in that it measures an average of squared deviations but it differs in that the deviation is taken with respect to zero rather than the distribution mean. (In statistical terms, variance is the second moment around the mean whereas ESD is the second moment around the origin.)

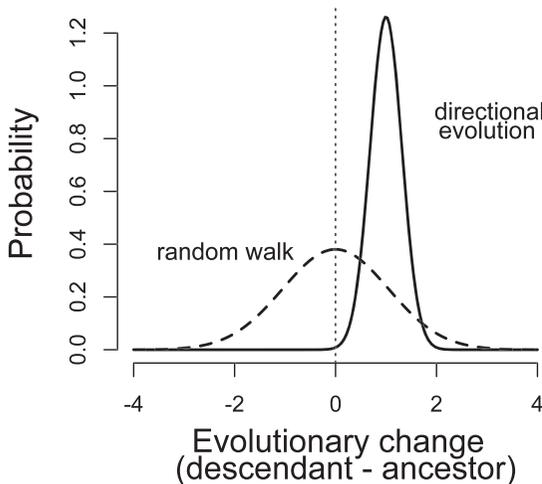


FIGURE 1. Distribution of evolutionary changes ( $\Delta X$ ) across ten generations for two models: random walk with  $\sigma_s^2 = 0.11$  (dashed line) and directional evolution with  $\mu_s = 0.1$  and  $\sigma_s^2 = 0.01$  (solid line). The expected change under directional evolution is nearly always positive, whereas the random walk is centered on zero. Despite this difference, these two parameterized models produce equal magnitudes of change over ten generations as measured by expected squared divergence (ESD = 1.1 for both).

For any probability distribution of evolutionary changes with mean  $M$  and variance  $V$ , ESD is equal to  $M^2 + V$  (Kendall and Stuart 1969: p. 58). These two terms represent evolutionary divergence from directional and nondirectional components, respectively. Because we already know the mean and variance of  $\Delta X$  for our three models of evolution, it is straightforward to compute the ESD of each as a function of elapsed time (Table 1).

### From Divergence to Rate Metrics

For any model, the form of the relationship between ESD and elapsed time will determine the most useful way to measure evolutionary rate. This is most straightforward for the random walk. Here, ESD increases linearly with time with a slope equal to the variance of the distribution of evolutionary steps,  $\sigma_s^2$  (Table 1). Thus, the magnitude of evolutionary divergence over any specified time interval is determined solely by  $\sigma_s^2$ , and this parameter is therefore a natural measure of evolutionary rate for a random walk. As  $\sigma_s^2$  increases, evolutionary trajectories experience larger positive and negative changes in trait value, resulting in net larger amounts of evolutionary divergence.

Rate is less straightforward for directional evolution because ESD increases as a quadratic function of time, depending on both the mean ( $\mu_s$ ) and variance ( $\sigma_s^2$ ) of the step distribution. When both the directional and nondirectional components of divergence are substantial, ESD is not readily reduced to a simple rate metric. However, if most evolutionary change is attributable to the directional component of change,  $\mu_s^2 t^2 \gg \sigma_s^2 t$ , and the latter term can be ignored. In this situation, the magnitude of divergence for a specified interval depends on  $\mu_s^2$ , which is therefore a useful rate metric. Conversely, if directionality is very weak, then most evolutionary change is caused by the variation around the (weak) trend, and  $\sigma_s^2$  will be an appropriate measure of rate. Clearly, the relative magnitudes of  $\mu_s^2$  and  $\sigma_s^2$  influence the degree to which the directional component of evolution dominates divergence. Perhaps less obviously, the duration of time also matters because

the trend term increases with the square of elapsed time, becoming increasingly important as the temporal window expands.

The ESD for the stasis model has two terms. The first,  $(\theta - X_A)^2$ , encompasses the attraction of the trait to the long-term mean, and the second,  $\omega$ , is the variance around that optimum. If the ancestral trait value is at or very near the optimum, the first term disappears and ESD is simply  $\omega$ . Alternatively, we can take the expectation over all ancestral trait values, in which case the expected value of  $(\theta - X_A)^2$  is the variance of the trait,  $\omega$ , and ESD is therefore  $2\omega$ . Either way, ESD depends only on  $\omega$  and so this parameter is a natural rate metric for this stasis model. Note that elapsed time does not appear in the ESD under stasis—we expect the same magnitude of change over one generation as over a million. This property is consistent with recent qualitative accounts of stasis on paleontological time scales (Gould 2002; Eldredge et al. 2005).

In addition to these real evolutionary differences, divergence in all empirical data sets will have an additional component from sampling error that can account for a surprisingly large proportion of the variation among samples (Hunt 2008a). Likelihood methods can estimate the model parameters in Table 1 in a way that accounts for this sampling noise (Hunt 2006).

All these rate metrics are squared terms because ESD is itself a squared quantity. If one desires rate metrics that are of the same dimensionality as the data (Gingerich 2009), the square root of these expressions may be used.

### Interpreting These Rate Metrics

The rate metrics listed in Table 1 may not look familiar, but each can be linked to an existing approach in evolutionary biology (Table 1).

*Directional Evolution Model* → *Traditional Rate Metric*.—The rate metric derived assuming strongly directional evolution,  $\mu_s^2$ , will approximate the square of traditional rate metrics. Assuming constant sampling error, the maximum likelihood estimate of  $\mu_s$  for an ancestor-descendant pair is  $\Delta X/t$  (Hunt 2006: p. 601), equivalent to the traditional rate

metric. For example, the compilation of trait sequences from Hunt (2007) analyzed below yields a correlation coefficient between  $\mu_s$  and traditional net rates that is greater than 0.99. Because of this correspondence, traditional rates can be justified when evolution is strongly directional.

*Random Walk Model*  $\rightarrow$  *Step Variance Metric*.—If traits evolve according to an unbiased random walk, the step variance parameter ( $\sigma_s^2$ ) is a sensible measure of evolutionary rate. This parameter has been occasionally employed as a measure of rate in paleontology (e.g., [Hunt and Carrano 2010](#)). Lynch's (1990) rate metric,  $\Delta$ , is a scaled version of the step variance, and it too has been applied on occasion to fossil data ([Cheetham et al. 1993, 1994; Geary et al. 2010](#)). This metric is equivalent to  $\sigma_s^2/(2V_P)$ , where  $\sigma_s^2$  is the step variance estimated with time measured in generations and  $V_P$  is the phenotypic variance of the trait ([Geary et al. 2010](#)). Lynch's  $\Delta$  has a known range under the scenario of mutation-drift equilibrium, approximately  $5 \times 10^{-5}$  to  $5 \times 10^{-3}$  ([Lynch 1990](#)), which provides a useful benchmark for judging the speed of evolutionary change when evolution is measured as  $\Delta$  or  $\sigma_s^2$ .

In contrast to its relative obscurity in the paleontological literature,  $\sigma_s^2$  is the standard rate metric used in comparative biology (e.g., [Felsenstein 1985; O'Meara et al. 2006; Ackerly 2009; Harmon et al. 2010](#)), although it is not known as such. This literature refers to Brownian motion (BM), rather than random walks, and thus the equivalent quantity is called the BM variance, instantaneous variance, or BM rate parameter, and it is often symbolized as  $\beta$ . BM and random walks are closely related models; as evolutionary steps decrease in magnitude and temporal spacing, in the limit, a random walk converges on Brownian motion. These models have similar properties, including expected changes that are normally distributed as described above.

The use of the step variance as a rate parameter is at odds with Bookstein's (1987) declaration that "random walks have no rates," but his statement is true only in a narrow, mathematical sense (see [Foote 1991; Hunt 2008a; Gingerich 2009](#)). As [Gingerich](#)

(2009) noted, [Bookstein](#) introduced an entity similar to the step variance called "reduced speed" which he treated as a rate parameter.

*Stasis Model*  $\rightarrow$  *Disparity Metric*.—Under stasis, a natural measure of divergence and rate is  $\omega$ , which is simply the total variance of all included samples with the contribution from sampling error subtracted ([Hunt 2006](#)). Effectively, this approach measures morphological change by its magnitude alone. This tactic is not often applied to anagenetic sequences, but it does represent a standard approach in studies of morphological disparity ([Roy and Foote 1997; Ciampaglio et al. 2001](#)) in which morphological variety is measured as the variance of forms, with no attempt to account for the time over which such differences accrued (but see [Brusatte et al. 2011](#)).

#### Rate Estimates from Simulated Data

Later in the paper, data compilations will be used to assess the behavior of these metrics in real empirical situations. But the first step for any method is to assess how well it performs under known conditions. In this section I assess the performance of the three rate metrics in Table 1 across different generating models of evolution, and across different temporal resolutions. Here I focus on three example scenarios: strongly directional evolution ( $\mu_s = 0.02$ ,  $\sigma_s^2 = 0.0002$ ; Fig. 2A), a random walk ( $\sigma_s^2 = 0.0002$ , Fig. 2B), and stasis ( $\omega = 1$ ; Fig. 2C). For each scenario, I generated 10,000 evolutionary sequences with 21 samples across resolutions spanning four orders of magnitude with 1, 10, 100, 1000, and 10,000 generations separating adjacent samples. All sequences were generated assuming within-sample variance of unity, and sample sizes of 30 individuals per sample; these two conditions determine the amount of sampling noise in the generated sequences. Directional evolution, random walk, and stasis models were fit via maximum likelihood to each sequence. The resulting parameter estimates were used to compute the rate metrics in Table 1. Simulations and model fitting were performed using the functions *sim.GRW*, *sim.Stasis*, and *fit3models* in the R package *paleoTS* ([Hunt 2011](#)).

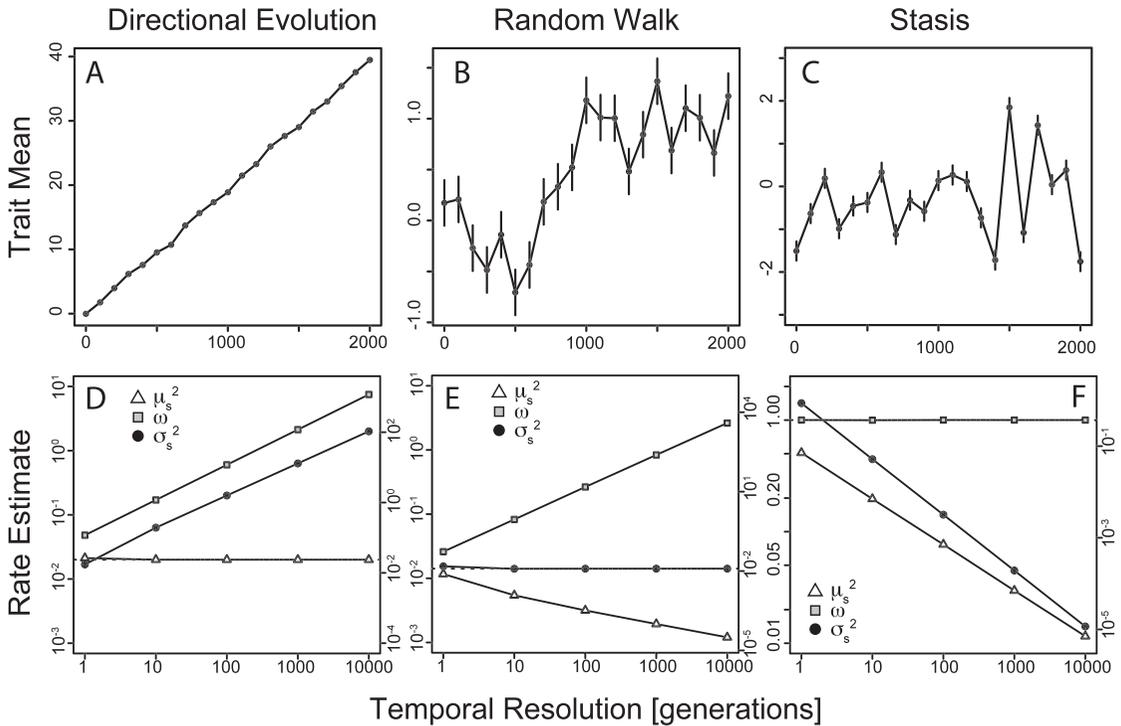


FIGURE 2. Examples of evolutionary sequences generated to assess the performance of the rate metrics in Table 1. Three models were considered: directional evolution (A), random walk (B), and stasis (C). Bottom row shows the temporal scaling of three rate metrics across directional evolution (D), random walk (E), and stasis (F) modes of change. The legend indicates the three rate metrics used, as defined in Table 1. For each evolutionary mode, the rate metric derived assuming that mode can be estimated well at any temporal resolution, but the other rate metrics always show systematic scaling with temporal interval. Dotted lines indicate true generating parameter values. Directional evolution (D) axes: left axis scale for  $\mu_s^2$  and  $\sigma_s^2$ , right axis for  $\omega$ . Random walk (E) and stasis (F) axes: left axis scale for  $\sigma_s^2$  and  $\omega$ , right axis for  $\mu_s^2$ .

Mean rate estimates across the three modes of evolution are plotted with respect to temporal resolution in Figure 2D–F. The main result is that estimated rates are uncorrelated with temporal resolution when their assumed model is true, but all show systematic temporal scaling otherwise. In other words, if evolution follows one of these three modes of evolution exactly, we can choose to measure evolutionary rates with the appropriate metric and thereby avoid the complications of temporal scaling. However, if we choose wrongly, rate estimates show systematic relationships with temporal resolution. One consequence of this is that the rate metrics that will be most useful are probably those whose assumptions most closely match realistic evolutionary patterns, a consideration that I will return to later in the paper.

Some aspects of these results have been reported before. For example, maximum likelihood estimates for the random walk

and directional evolution models were previously shown to be unbiased, or nearly so, regardless of temporal resolution or the strength of the directional trend (Hunt 2006). Gingerich (1993) showed that traditional rate metrics are uncorrelated with interval length under strongly directional change, and that they have steep negative slopes under stasis and intermediate negative slopes under random walks. The relevant rate metric,  $\mu_s^2$ , shows similar relationships (open triangles in Fig. 2D–F).

The fact that  $\mu_s^2$  and  $\sigma_s^2$  can be estimated fairly across a range of temporal resolutions has implications for one commonly advanced interpretation of the empirical negative scaling of rate with time first reported by Gingerich (1983). Gould (1984) argued that this negative relationship was an inevitable mathematical artifact of plotting a ratio (evolutionary rate) against its denominator (elapsed time). This argument was subsequently explored further,

most notably by Sheets and Mitchell (2001a) and Roopnarine (2003). The maximum likelihood estimators of  $\mu^2_s$  and  $\sigma^2_s$  both have elapsed time in the denominator (Hunt 2006: p. 584), and so the curves for these metrics effectively plot a ratio versus its denominator. Nevertheless, these metrics show no systematic scaling with temporal resolution when their underlying models are true and so the negative scaling of rates with respect to time cannot be a mathematical necessity. Knowing the true model of evolution essentially gives the correct temporal scaling of divergence over time, which in turn allows for estimation of rate, regardless of temporal interval. Thus correlations between a rate metric and interval length tell us something about how evolutionary dynamics differ from the model underpinning that rate metric.

#### Summary: The Relationship between Tempo and Mode

Evolutionary tempo can be measured with model parameters that relate evolutionary divergence to elapsed time. For each of the three canonical modes of evolution—directional change, random walk, and stasis—a reasonable rate metric can be derived and related to existing strategies in evolutionary paleobiology. Simulations indicate that we can accurately infer these metrics via likelihood at any temporal resolution if the underlying mode of change is simple and known, but that this convenient behavior is not likely when real evolutionary dynamics deviate from the assumed model.

#### LRI and Generational Rates

The approach implemented in the previous section starts with simple models and works forward to determine sensible ways of measuring rate, given those models. Another method has been proposed to decompose tempo and mode: the log-rate, log-interval (LRI) method (Gingerich 1993, 2001, 2009). This approach produces estimates of evolutionary change on the time scale of a single generation, called intrinsic, base, or generational rates. I did not use this terminology above, but likelihood estimates of model parameters are also “generational” in the

sense that they accurately measure evolutionary changes over a single time step when the assumed model is correct. Unlike these likelihood-based rates, generational rates from LRI are said to be valid regardless of the underlying mode of evolution. If so, their use obviates the complicated dependence of rate on mode explored in the previous section, and this would lend LRI rates an important advantage over likelihood-based estimates.

The original log-rate, log-interval plot from [Gingerich \(1983\)](#) generated substantial debate about its implications for the relationship between micro- and macroevolution (Gould 1984; Stanley 1985; Sheets and Mitchell 2001a; Gingerich 2009). My goal in this section is narrower and more practical: to assess if generational rates produced by the LRI method usefully measure the pace of evolutionary change under realistic conditions.

#### The LRI Procedure

Gingerich developed LRI to determine the mode of evolution simultaneously with the rate of change over a single generation (Gingerich 1993, 2001, 2009). This approach computes all pairwise rates of change among samples in an evolutionary sequence and plots the logarithm of these rates as a function of the logarithm of their corresponding interval lengths. Time is measured in generations and rates in haldanes; by convention the logarithms are base 10.

The expected slope of this log-rate, log-interval plot depends on the mode of evolution. Strictly directional evolution results in a slope of zero, stasis a slope of  $-1.0$ , and random walks produce intermediate slopes of  $-0.5$  (Gingerich 1993; see Sheets and Mitchell 2001b about the variation around these expectations). The intercept of this relationship is the predicted change at a  $\log(\text{interval})$  of zero, which is a single generation. Taking the antilog of the intercept converts it to a rate over one generation on the original scale. Gingerich (1993) recommends fitting this line using a robust algorithm that minimizes the sum of absolute deviations. Examples of the LRI approach applied to two aspects of molar

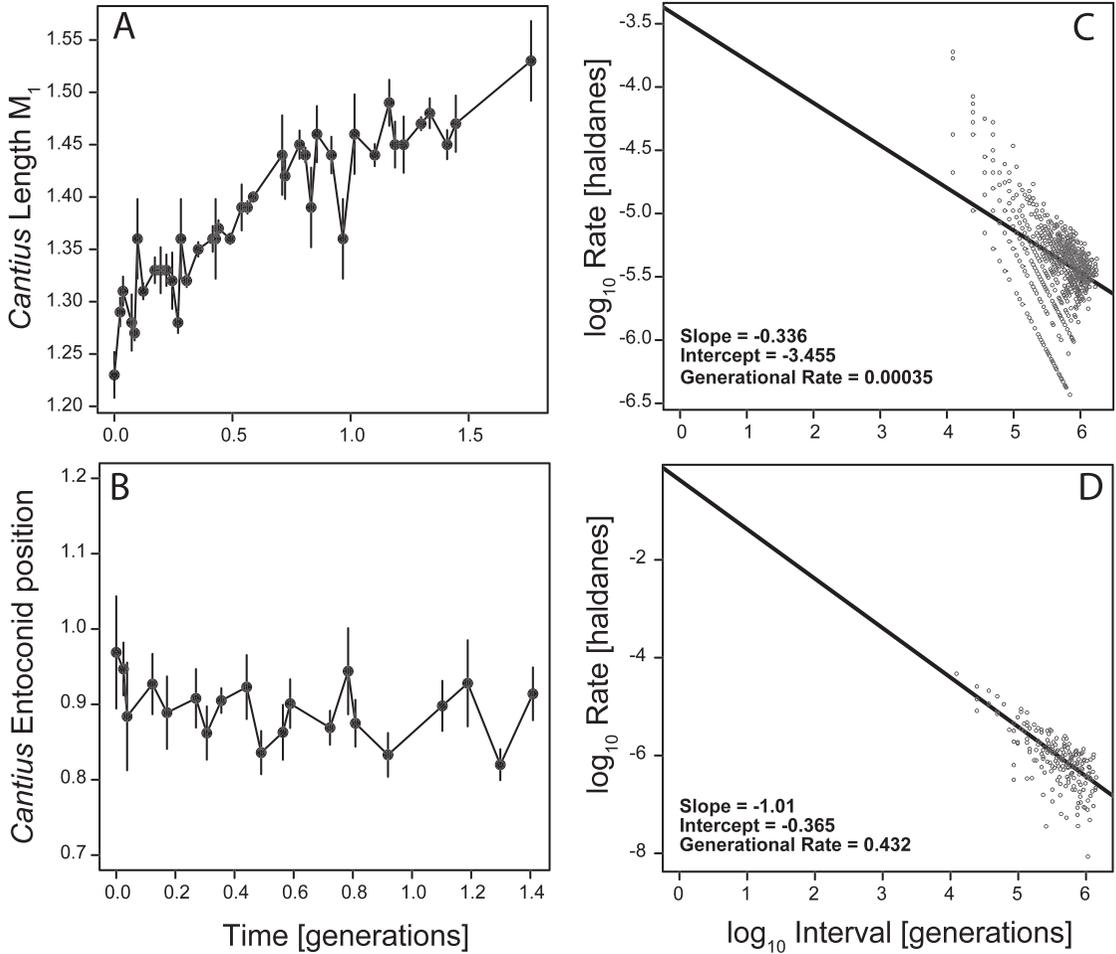


FIGURE 3. Evolutionary sequences showing, for the lower first molar of the primate *Cantius*, mean length (A) and position of the entoconid cusp ( $x$ -shape coordinate) (B). Error bars indicate one standard error. C and D, LRI plots for the same evolutionary sequences; calculations and plot done by the *LRI()* function in the *paleoTS* package. Data from Clyde and Gingerich (1994).

morphology in the primate lineage *Cantius* are given in Figure 3 (data from Clyde and Gingerich 1994).

#### Concerns about LRI Generational Rates

Despite their intuitive appeal, several lines of reasoning suggest caution about LRI generational rates. First, for data at typical paleontological resolutions, roughly  $10^4$ – $10^6$  generations between adjacent samples, the intercept from which generational rates are computed lies far outside the range of observed intervals (Fig. 3C,D). As with any extrapolation, small errors in slope can propagate into very large errors in the intercept, especially on a log-log scale.

Second, LRI sometimes produces puzzlingly slow generational rates for trait sequences with substantial evolutionary change. For example, compare the two evolutionary trajectories shown for *Cantius* (Fig. 3A,B). Through this interval, length of the lower first molar changes dramatically, increasing by nearly eight standard deviations. In contrast, the position of the entoconid cusp varies little through the same sequence. In fact, the observed variation among these samples is no greater than that expected from sampling error alone; the maximum likelihood estimate for  $\omega$  of the stasis model is zero. Apparently, no detectable evolutionary change whatsoever has occurred in this trait,

yet its LRI generational rate is over 1200 times greater than that of molar length (0.432 versus 0.00035; Fig. 3C,D). It is difficult to imagine how even the direction of this contrast could be correct. It is conceivable that each generation experienced large fluctuations in entonid position that averaged out exactly to produce no differences at the observed temporal resolution. But if so, where could this information possibly reside, given that the observed samples show no evidence for real differences?

Third, most estimates of generational rates fall in a narrow range between 0.1 and 0.3 standard deviations per generation (Polly 2002; Gingerich 2009). These values indicate rather large generational change, even more substantial than one would predict from measurements of natural selection in living populations. The expected evolutionary response to selection (in trait standard deviation units) is equal to the standardized selection coefficient multiplied by trait heritability (Lande and Arnold 1983). Taking 0.4 as a typical heritability of morphological traits (Mousseau and Roff 1987) and 0.16 as a typical selection coefficient (Kingsolver et al. 2001) yields an expected rate of 0.064 haldanes over a single generation. Thus, LRI generational rates estimated in paleontological studies are routinely two to five times higher than these predicted rates, even though the compilation of selection coefficients should be systematically biased in favor of strong selection (Kingsolver et al. 2001; Hersch and Phillips 2004; Kingsolver and Pfennig 2007; but see Knapczyk and Conner 2007).

Finally, observed rates should, as a general rule, overestimate true rates because of sampling error. Paleontological samples are obviously finite, and so some proportion of the observed differences between samples must be attributable to noise in estimating sample means (Hendry and Kinnison 1999; Kinnison and Hendry 2001; Hunt 2006). Perhaps most troubling about LRI rates is that the typical range for generational rates matches almost exactly the expectation from sampling error alone. In the absence of any

real evolutionary change, sampling noise will produce apparent rates of change from one generation to the next that average  $2/\sqrt{n\pi}$  haldanes, where  $n$  is the number of individuals measured in each generation (Appendix 1). Average sample sizes in paleontological sequences typically range from about  $n = 10$  to  $n = 50$ , which translate to generational rates from 0.16 to 0.36, very nearly matching the typical empirical range for LRI rates. Thus, even if true evolutionary rates are much lower, the contribution from sampling error should tend to inflate estimates into the empirically observed range (see also the similar result obtained via simulation by Kinnison and Hendry [2001]).

### Testing LRI via Simulation

*Simulation Procedure.*—In light of these concerns, I assessed the performance of LRI generational rates via a series of simulations. All involved a lineage sampled 15 times with  $10^5$  generations between each sample. Traits were evolved as a random walk with step variances ranging from  $10^{-9}$  to  $10^{-1}$  in order-of-magnitude increments. Within-sample trait variance was set to unity, and therefore these step variances span rates of change from those much slower than drift to those faster than drift (neutral prediction:  $10^{-4} < \sigma_s^2/V_P < 10^{-2}$ ). Each sequence was simulated assuming two different sample sizes for each population:  $n = 5$  and  $n = 50$ , with 10,000 replicates for each combination of sample size and step variance. Simulations were completed using functions from the *paleoTS* package (Hunt 2011); the complete R code is provided in Supplementary Appendix 1.

The step variance measures variance in the distribution of evolutionary increments, but LRI generational rates reflect a different measure of dispersion for this distribution, its mean absolute deviation (Gingerich 2009: p. 665). When the distribution of steps is Gaussian (as in these simulations),  $\sigma_s^2$  can be computed as the squared LRI generational rate multiplied by  $2/\pi$  (see Appendix 1). This calculation was performed to convert the generational rates estimated by LRI to step variances for comparison to the generating parameter of the simulations.

*LRI Simulation Results.*—The plot of the LRI estimated rates versus the true simulated rates reveals several important features (Fig. 4, bottom panel). When true rates are very high, the LRI rates are approximately unbiased—they closely follow the line of equality on the graph. However, when true rates are low, LRI estimates are severely biased upward, often by several orders of magnitude. Moreover, when rates are low to moderate, inferred generational rates are inversely correlated with true rates: as the pace of evolutionary change speeds up, rates estimated by LRI actually decrease. LRI rates are also quite sensitive to the amount of sampling noise. For the same true rate of change, estimated LRI rates for  $n = 5$  can be orders of magnitude higher than when  $n = 50$ . The distribution of empirical, likelihood-based step variance estimates (Fig. 4) suggests that true paleontological rates will typically fall on the slower end of this figure, in the region for which LRI estimates are biased and negatively correlated with true rates of change.

These patterns are all consistent with sampling error confounding estimation of LRI generational rates. When rates are fast enough that true evolutionary differences are much larger than sampling noise, LRI rates are unbiased. However, as true rates decrease, the proportional contribution of sampling error increases (Fig. 4, top panel). This noise is mistaken for true evolutionary divergence, causing an upward bias in LRI rates. Moreover, because sampling error is analytically equivalent to stasis (Hannisdal 2006; Hunt 2006), noise causes the LRI slope to steepen toward stasis-like values. Steepening the slope elevates the intercept, further inflating generational rates. At the lower extreme of rates very close to zero, the LRI rates converge close to the expectation from sampling noise derived above (Fig. 4).

How large does sampling error need to be to compromise LRI estimates of rates? The top panel in Figure 4 plots the proportion of the variance between consecutive samples that is attributable to sampling error, separately for  $n = 5$  and  $n = 50$ . Evolutionary variance for a random walk is  $\sigma^2_s t$  (Table 1) and sampling

variance is  $2V_P/n$  (Hunt 2006); this graph plots the latter divided by the sum of these two components. The LRI generational rates converge to nearly unbiased estimators only when sampling variance accounts for less than about 5% of the total variance between consecutive samples (Fig. 4). Because empirical paleontological sequences usually have levels of sampling noise higher than this threshold (Hunt 2008a), LRI generational rates will often be strongly affected by sampling error.

These effects of sampling error would also seem to explain the counterintuitive finding of very low generational rates in the highly trended molar length sequence from *Cantius* (Fig. 3). Because evolutionary differences are very large relative to sampling noise here, sampling error only slightly inflates LRI rate estimates. This interpretation suggests that the anomalously low generational rates occasionally produced by LRI are probably closer to the true rates than are those in the typical range of 0.1–0.3, which mostly reflect sampling noise.

At high rates of evolution, both LRI and maximum likelihood approaches are approximately unbiased, but maximum likelihood rates are far more precise. Figure 4 shows the middle 50% of rate estimates (interquartile range) as error bars for the LRI estimates for  $n = 50$ . The comparable range for maximum likelihood estimates is shown as a gray envelope. Over the range of rates for which both LRI and likelihood rates are unbiased, the maximum likelihood estimates are orders of magnitude more precise.

### Are Generational Rates Always Accessible to Paleontologists?

Generational rates may be poorly estimated by LRI, but simulations presented above demonstrate that they may be recovered by maximum likelihood estimation, even from coarsely sampled sequences. This is perhaps surprising—we are translating patterns on geological time scales into inferences at generational resolutions. Is this translation always possible? In this section, I show that this question must be answered in the negative. Some realistic models of evolution produce evolutionary trajectories for which

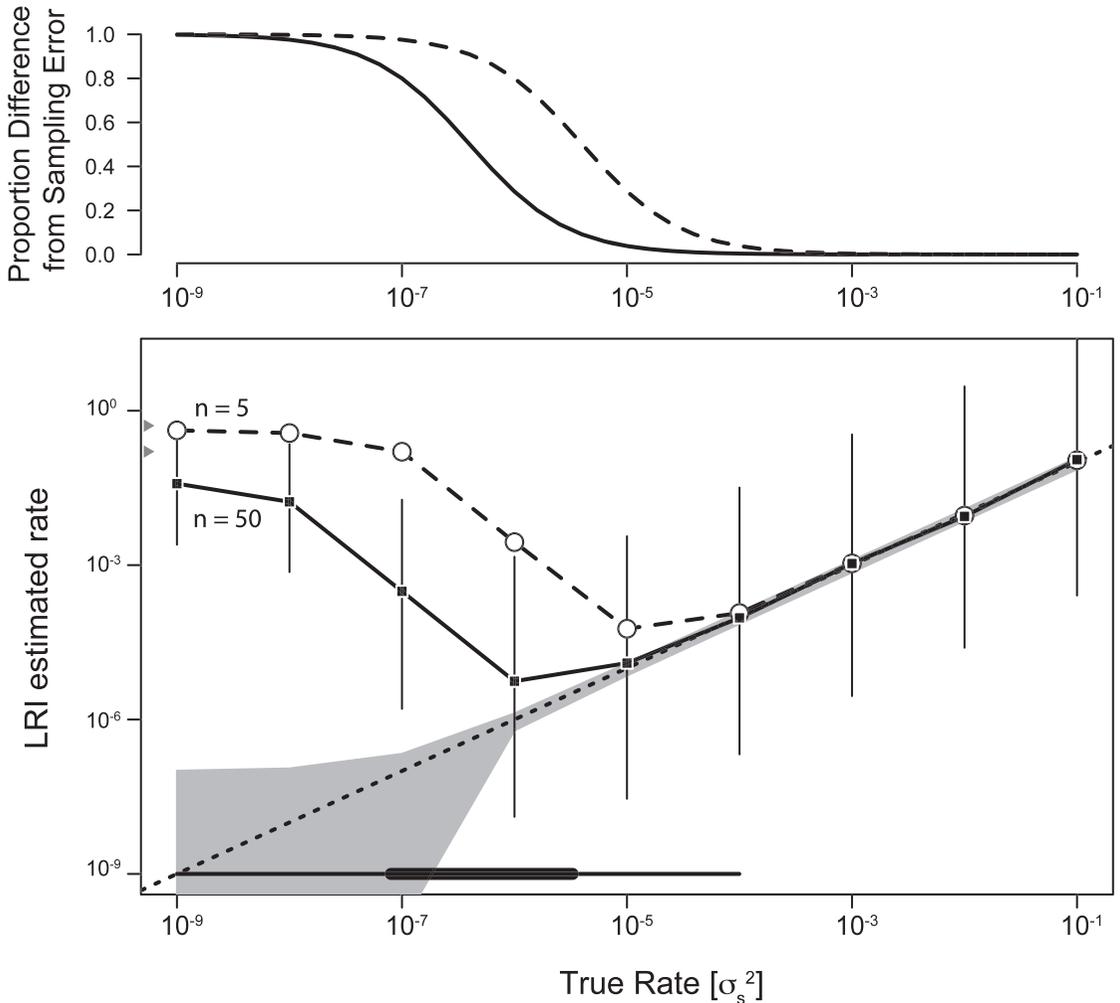


FIGURE 4. Performance of LRI method for estimating generational rate in simulated sequences. Bottom panel plots the true, simulated rate against the median estimated LRI rate across 10,000 simulations; dotted line is the line of equality. Simulations were performed at two levels of sampling noise:  $n = 5$  individuals (dashed line, open circles) and  $n = 50$  individuals (solid line, filled squares). The precision of the estimates is presented as the interquartile range; for clarity these are presented only for  $n = 50$ . The gray envelope shows the equivalent precision for maximum likelihood estimates from the same simulations. Small gray triangles at the far left show the expected rate from sampling error alone (top triangle is  $n = 5$ ; bottom one is  $n = 50$ ). The horizontal bars at the bottom indicate the 90% (thin line) and 50% (thick line) quantiles of the empirical distribution of paleontological step variances from Figure 7B. Top panel shows for these same simulations the proportion of the variance between consecutive samples that is attributable to sampling error (dashed line is  $n = 5$ ; solid line is  $n = 50$ ). At the lowest rates considered, essentially all differences are sampling noise; at the fastest rates, sampling error is negligible.

generation-to-generation change is simply inaccessible when sampled at coarse temporal resolution. No analysis, no matter how sophisticated, can convert paleontological divergence into generational rates under these scenarios.

In this section, I focus on a model of a population evolving under the influence of a nearby peak in the adaptive landscape (Lande 1976). When the population is displaced from

the optimum, it evolves toward the optimal trait value, rapidly at first, but slowing over time to trace an exponential approach (Fig. 5). This is the expected adaptive trajectory when a population invades a new environment that differs from its ancestral habitat or when the environment changes suddenly. This model is an Ornstein-Uhlenbeck (OU) process, and it has four parameters that determine the evolutionary trajectory: the initial ( $X_0$ ) and

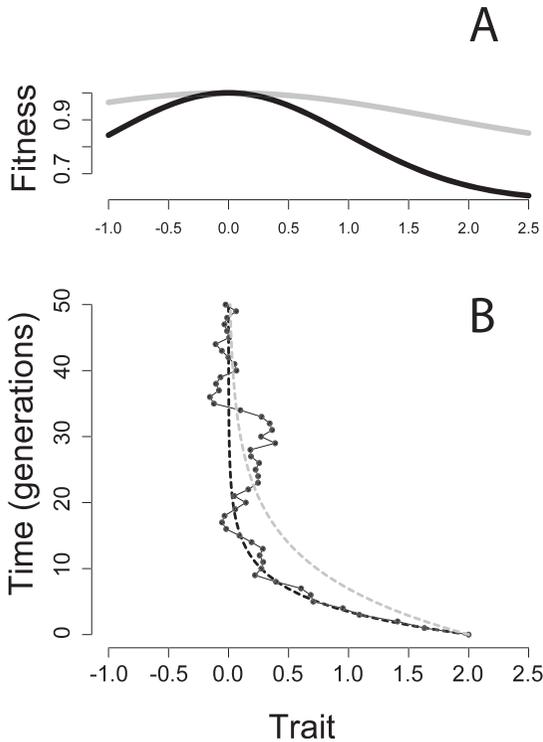


FIGURE 5. Evolution of a population with a nearby peak in the adaptive landscape (Ornstein-Uhlenbeck process). A, Two adaptive landscapes with trait optima at zero. Stabilizing selection is stronger for the black curve than the gray curve because fitness drops off more sharply away from the optimum in the former. B, One realized trajectory of a population (filled black symbols) evolving on the black adaptive landscape from panel A. Dotted black and gray lines show the average trajectory for the corresponding adaptive landscapes in A.

optimal ( $\theta$ ) trait values, a parameter ( $\alpha$ ) that determines the strength with which the optimum attracts the population, and a step variance ( $\sigma_s^2$ ) that encompasses the stochastic input from genetic drift (this parameterization follows Hansen 1997; see also Arnold et al. 2001). The expected (or average) trajectory is determined by the difference between  $X_0$  and  $\theta$  and by the strength of natural selection. High values of  $\alpha$  indicate steep declines in fitness away from the optimum and correspondingly rapid exponential approaches; weaker selection results in shallower, but still exponential, trajectories (the fitness surface is assumed to be Gaussian; Fig. 5A).

The mean and variance of the distribution of evolutionary changes under this model are

known and so it is possible to write out an expression indicating the expected squared divergence (ESD) for this adaptive model over any specified interval length. This expression is described in Appendix 2. I do not include it here because it is a complex function of all four model parameters and elapsed time, and no simple rate metric is suggested by its form.

Although cumbersome, the expression for ESD of this model can be used to explore the expected magnitude of change under this model at any temporal scale. The base parameters for this exploration were:  $\alpha = 10^{-3}$ ,  $\sigma_s^2 = 10^{-2}$ , and  $X_0 = \theta = 0$ . Starting the trait value at the optimum omits the initial exponential adaptive phase, leaving only the trait fluctuations near the adaptive peak; this choice also simplifies considerably the expression for ESD (Appendix 2). Divergence, measured as ESD, was plotted as a function of elapsed time under four scenarios: (i)  $\alpha$  and  $\sigma_s^2$  as above, (ii)  $\alpha$  multiplied by five, (iii)  $\sigma_s^2$  multiplied by five, and (iv)  $\alpha$  and  $\sigma_s^2$  each multiplied by five.

The shape of divergence versus temporal resolution is the same for all parameter combinations: ESD first rises and then flattens as elapsed time between samples increases (Fig. 6). The initial rise is proportional to the step variance ( $\sigma_s^2$ ) like a random walk but becomes stasis-like over the long term with an asymptote that depends on the ratio of  $\sigma_s^2$  to  $\alpha$  (Appendix 2). Biologically, this means that the magnitude of trait change near the optimum over a single generation depends only on genetic drift, but changes over longer intervals reflect a balance between wandering induced by genetic drift and selection pulling the trait value back toward the optimum. As a result, divergence shows qualitatively different behavior at generational time scales compared to the much coarser resolutions usually available to paleontologists.

It is this difference in behavior between fine and coarse resolutions that makes it impossible to predict short-term, generational divergence from divergences occurring over much longer intervals. Consider curves (i) and (iv) in Figure 6. Increasing  $\sigma_s^2$  has the effect of elevating generational change under scenario

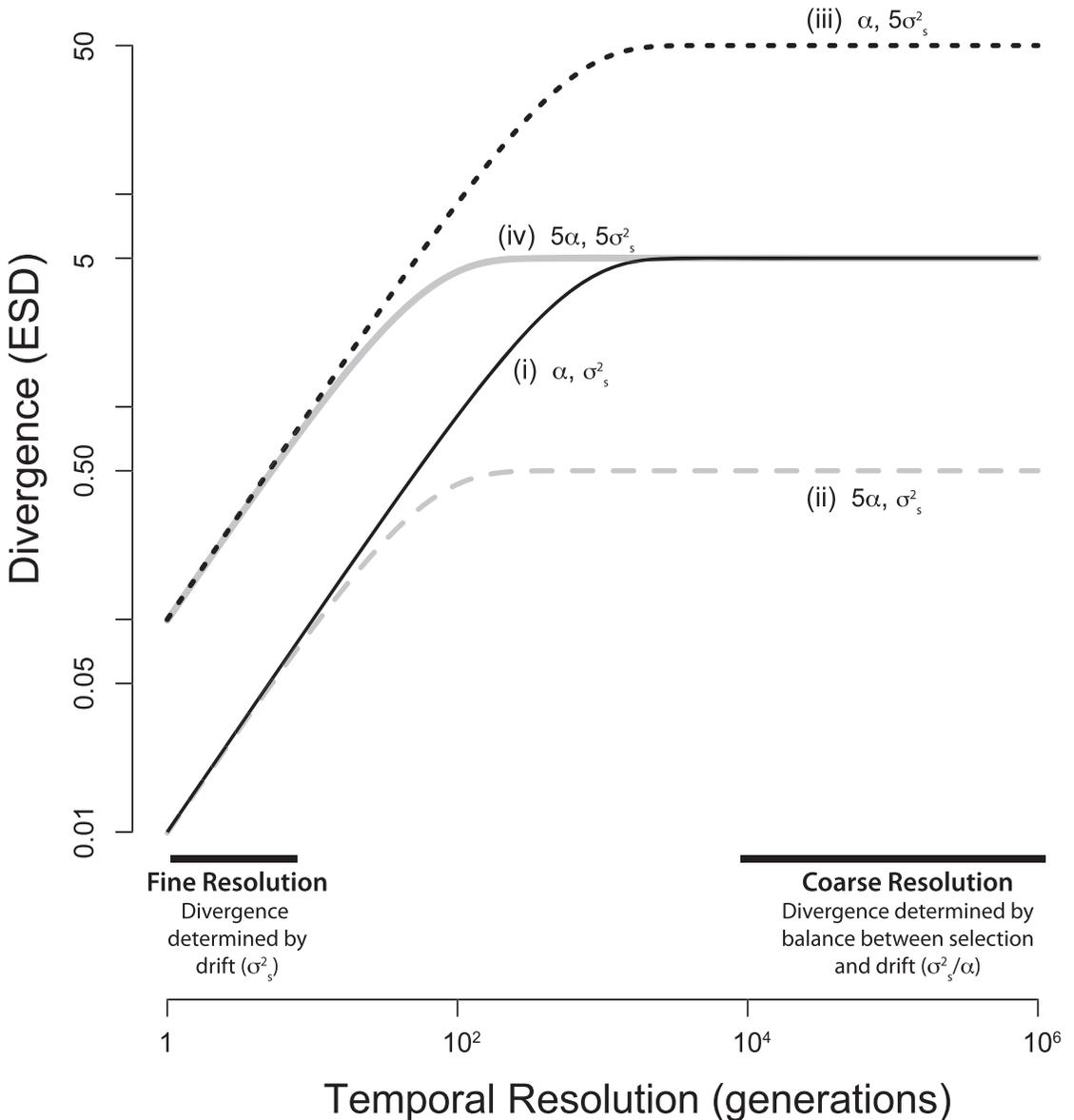


FIGURE 6. Magnitudes of expected change (expected squared divergence, ESD) with respect to elapsed time for a population starting out at a peak on the adaptive landscape. The four curves show four different combinations of parameter values: (i)  $\alpha = 0.001$ ,  $\sigma_s^2 = 0.01$ ; (ii)  $\alpha = 0.005$ ,  $\sigma_s^2 = 0.01$ ; (iii)  $\alpha = 0.001$ ,  $\sigma_s^2 = 0.05$ ; (iv)  $\alpha = 0.005$ ,  $\sigma_s^2 = 0.05$ . Over short periods of time, divergence is determined by drift alone, and increases nearly linearly with the step variance. Divergence over time eventually plateaus at a value determined by a balance between natural selection and genetic drift.

(iv), but when sampling is coarser than every  $\sim 1000$  generations the two scenarios produce exactly the same average divergence and thus they will be paleontologically indistinguishable despite their difference in generational rate. Conversely, other combinations of parameter values have the same generational rate, but result in very different divergences at geolog-

ical resolutions (compare scenarios i versus ii and iii versus iv).

This simple adaptive model differs from random walks, stasis and directional change in that those models imply a clear rate metric that holds over all temporal scales. Those three models are also unusual in that they imply uniform dynamics—one expects the

same behavior at all times, and for all trait values. It seems probable that real evolutionary dynamics are not so homogeneous, and that most heterogeneous evolutionary processes will not imply a simple and resolution-independent rate metric. If we generalize the adaptive model above to allow for the movement of peaks on the adaptive landscape in response to secular changes in the biotic or abiotic environment, evolutionary dynamics still differ over short and long time intervals. Over a few generations, population changes are dominated by the ascent of local peaks in the adaptive landscape, which is generally quite fast. Over much longer periods of time, divergence is controlled by how the peaks themselves move in response to changing selective conditions (Arnold et al. 2001; Estes and Arnold 2007). For these scenarios that imply heterogeneity, short-term, generational changes cannot be predicted from observations made at coarser resolutions.

This section injects two additional notes of caution into this exploration of evolutionary rates. First, not all models of evolution imply a sensible rate metric. Expected magnitudes of change can be a complex function of multiple parameters, none of which are easily interpretable as a rate parameter. Second, for some—possibly most—realistic notions of evolutionary change, it is simply impossible to translate observations at coarse temporal resolutions into estimates of change from one generation to the next.

### Empirical Rates from Fossil Lineages and Extant Clades

The preceding sections consider the behavior of different rate metrics under a variety of modeled conditions. The results help to clarify the choices available for measuring rates of phenotypic evolution, but only to a certain extent. In the most favorable situation, evolutionary dynamics are adequately accounted for as a random walk, stasis, or directional change, and the appropriate rate metric can be used to measure rates of change on any temporal scale. Still, two important complications remain. First, different styles of evolutionary change can operate across dif-

ferent lineages and traits (Erwin and Anstey 1995; Hunt 2007), and thus different data sets will imply different optimal rate metrics. But different metrics cannot be compared, which negates the primary reason for computing rates in the first place (Roopnarine 2003). Second, even when random walks, stasis, or directional evolution models *approximate* data well, they are seldom strictly true. And, as discussed in the previous section, more complex and realistic notions of evolutionary change are not likely to imply any rate metric that holds over all temporal resolutions and conditions.

Real data sets will often be heterogeneous and messy, and therefore the behavior of rate metrics derived from simple models should be evaluated empirically. In this section, I use recent compilations of morphological evolution in fossil lineages (Hunt 2007) and extant clades (Harmon et al. 2010) to document the behavior of the three rate metrics derived earlier in the paper. I consider the strength of temporal scaling as a criterion to evaluate rate metrics. To the extent that these rate metrics show weak correlations with interval length, rate estimates across different studies can be compared without concerns that differences in resolution are driving the results.

### Paleontological Data and Analysis

Each of the three rate metrics under consideration was estimated for the 251 sequences of trait values compiled in Hunt (2007). These time series include data on size, shape, and other phenotypic characteristics from 53 fossil lineages ( $n = 6\text{--}114$  samples per lineage). The relevant rate parameters were estimated via maximum likelihood using the functions in the R package *paleoTS* (Hunt 2011), using the joint parameterization (Hunt 2008a: p. 119) with the *fit3models* function. Time was measured in years and then separately in generations; generation lengths followed the original studies or were derived from information on living relatives (Supplementary Table 1). Many kinds of traits are represented, measured in different units, and so the raw parameter estimates are not comparable across sequences. To place them

on a common scale, rate parameters for each sequence were divided by the pooled, within-sample variance for that sequence (Hunt 2006).

### Phylogenetic Data and Analysis

Harmon et al. (2010) analyzed body size in a set of 49 clades of extant species with phylogenies scaled to geological time. For each of these clades, I fit the random walk (Brownian motion) and stasis models via likelihood using the function *fitContinuous* in the R package *geiger* (Harmon et al. 2009), setting this function's model argument to "BM" and "white," respectively, and accounting for sampling error as per Harmon et al. (2010). This procedure fits models across an entire clade assuming that both the models and model parameters are uniform across the entire tree. The resulting parameter from the Brownian motion model is  $\sigma_s^2$ , and the parameter estimate for the stasis model is  $\omega$  (see Hunt and Carrano 2010 on the interpretation of these models as applied to lineages and clades). A model of directional evolution cannot be fit to these data because in each tree all terminal taxa are equidistant from the root. Trends manifest as a systematic relationship between trait values and elapsed time and therefore their estimation requires variation in the ages of terminal taxa. The resulting rates were scaled as for paleontological metrics by dividing by within-population phenotypic variance using the composite body size variance calculated by Harmon et al. (2010: p. 2389).

Each case, whether fossil lineage or extant clade, yields rate estimates over some interval of time. To explore temporal scaling of these rates, it is necessary to decide which measure of time is most suitable. With fossil lineages, one could look at total sequence duration or temporal resolution measured as the mean time interval between consecutive samples. Clade duration (time elapsed since root), cumulative duration of evolutionary time represented (summed branch lengths), or any measure of resolution (e.g., mean branch length, mean temporal distance to nearest relative) are all possible measures of time for the comparative data. None of these choices is

obviously more appropriate than the others. For what follows, interval length is measured as the duration of the sampled lineage or clade. Using other metrics yields qualitatively similar results, the data for which are provided in Supplementary Tables 1 and 2.

### Temporal Scaling: Results and Implications

All rate metrics vary systematically with temporal resolution in at least one of the two data sets (Table 2, Fig. 7). In the fossil data, the rate metric assuming directional change ( $\mu_s^2$ ) shows a strong negative correlation with time interval measured in generations (Spearman rank correlation,  $r_s = -0.707$ ), the random walk step variance shows a weaker, negative scaling ( $r_s = -0.409$ ), and disparity has an even weaker but positive relationship with interval length ( $r_s = +0.214$ ). Among the phylogenetic data sets, step variance has a weak negative systematic relationship with time ( $r_s = -0.271$ ) but disparity increases moderately strongly with interval length ( $r_s = +0.412$ ). Relationships are quite similar if time is measured in years, rather than generations (Table 2).

The negative correlation between temporal resolution and rates in haldanes (Gingerich 1983; see also Fig. 7A) is not surprising because any evolutionary pattern other than strictly directional trends will result in such a pattern (Roopnarine 2003). The negative temporal scaling of step variance is less inevitable, however, as most of the suggested explanations for this relationship do not apply to this metric. Evolutionary reversals are inherent to the random walk model and they should not in themselves cause temporal scaling of the step variance. Indeed, the simulations presented above demonstrate that no temporal scaling is expected for traits that truly evolve according to this model.

Beyond evolutionary reversals, two additional causes of temporal scaling for the step variance metric should be considered. First, it has been suggested that the upper right portion of a rate-interval plot is unlikely to be populated because high rates sustained over long intervals lead to unrealistically large amounts of change (Gingerich 1983; Bush 2000; Sheets and Mitchell 2001a). This

TABLE 2. Rank correlations between the rate metrics and total temporal span for 251 trait sequences.  $r_s$  = Spearman rank-correlation coefficient. Temporal span is the duration of the fossil sequence or clade. As explained in the text, it is not possible to fit models of directional evolution to clades of entirely extant species.

Data	Rate metric	Time in generations	Time in Myr
Fossil lineages	$\mu_s^2$	$r_s = -0.707, p < 2.2e-16$	$r_s = -0.720, p < 2.2e-16$
	$\sigma_s^2$	$r_s = -0.409, p = 1.5e-11$	$r_s = -0.379, p = 5.5e-10$
	$\omega$	$r_s = +0.214, p = 0.0006$	$r_s = +0.105, p = 0.10$
Extant clades	$\mu_s^2$	—	—
	$\sigma_s^2$	$r_s = -0.271, p = 0.060$	$r_s = -0.156, p = 0.28$
	$\omega$	$r_s = +0.412, p = 0.003$	$r_s = +0.536, p = 7.2e-5$

effect is much reduced for the step variance as a metric compared to traditional rates. Although within-lineage step variances are absent in the upper right portion of Figure 7B, the presence of within-clade rates in this area, coupled with the observation that such rates are no faster than those implied by neutral genetic drift, indicates that such high-rate, high-interval observations are not biologically impossible.

An opposite effect has been posited for the lower-left portion of rate-interval plots. Very slow rates over short intervals involve morphological differences so small that they require astronomically large sample sizes to document (Gingerich 1983; Bush 2000; Sheets and Mitchell 2001a). Such small differences are normally swamped by sampling noise, placing a practical minimum on the detection of morphological difference. This effect is also much reduced when measuring rates as the step variance because this metric accounts for sampling error. When rates are truly very low, one ends up with many zero rates with a few estimates that are rather higher, but the average rate is still unbiased (Hunt 2006: p. 587). Because the relatively high rates observed over short intervals in Figure 7B are not accompanied by many zero-rate estimates, the negative temporal scaling in this plot is not driven by the difficulty of detecting very slow rates of change. (Zero rates are rarely obtained with traditional rates because a numerical coincidence is necessary to produce a morphological difference of exactly zero.)

Eliminating these two artifactual explanations supports the conclusion that the negative scaling of the step variance metric is caused by systematic deviation from the model of a random walk. The direction of

this scaling indicates less divergence over long periods than would be predicted by the magnitude of shorter-term changes. Evidently, evolution is often more bounded than a random walk—morphology can meander freely but only over a certain range, beyond which limits to change come into effect. This boundedness can derive from constraints or from adaptive limits inherent to lineages or clades (adaptive zones *sensu* Simpson 1944). The frequent support for stasis within lineages (Hunt 2007), coupled with the temporal scaling of the step variance within lineages, suggests that this boundedness is stronger within lineages than clades.

The low correlation between divergence (measured here as  $\omega$ ) and time within fossil lineages reported here and elsewhere (Sheets and Mitchell 2001a; Uyeda et al. 2011) is consistent with stasis as a common evolutionary mode within species. However, caution is required because the limits to divergence within species may be exaggerated by how species are recognized in the fossil record. Paleontologists may hesitate to assign populations to the same species if they are separated by large shifts in morphology, even if they are part of a single, unbranching lineage (Gingerich 1983, 1993).

For the same temporal window, rates are generally higher in the phylogenetic than in the fossil data sets (Fig. 7B,C). This offset may be interpreted as supporting the punctuated equilibrium model of change (Eldredge and Gould 1972): divergence in the extant clades includes intervals of higher rates associated with cladogenesis, whereas the paleontological rates incorporate only slower, anagenetic rates. This is a plausible explanation, but caution is called for because these data are very heterogeneous. Moreover, the relative

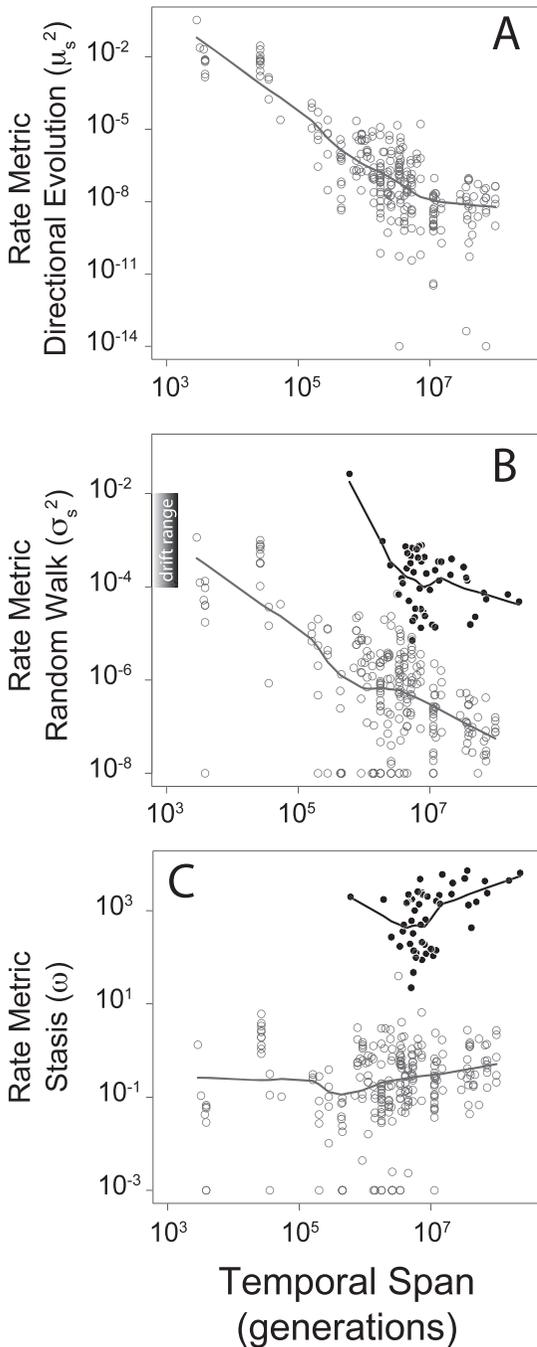


FIGURE 7. Temporal scaling of three different rate metrics. Shown are the rate metrics implied by strongly directional evolution (A), random walks (B), and stasis (C). Data are rates estimated from fossil lineages (open circles) and extant clades (filled circles). Curves are locally weighted regressions (lowess), which are useful for drawing curves through points without specifying a functional form. Estimates at or very close to zero are plotted at  $10^{-14}$  (panel A),  $10^{-8}$  (panel B), or  $10^{-3}$  (panel C).

positions of the anagenetic and phylogenetic clouds of points shift depending on how one chooses to measure elapsed time. A recent, parallel compilation of divergences over time using different measures of divergence and time found greater continuity between within-lineage and clade-level patterns (Uyeda et al. 2011), while still finding evidence that short-term, within-lineage evolution is more bounded than long-term, clade-level patterns.

## Discussion

### Rates As Model Parameters

Defining evolutionary rates in the traditional way as  $\Delta X/t$  may seem self-evident and therefore assumption-free. In agreement with Bookstein (1987) and Sheets and Mitchell (2001a), I have argued here that this intuitive notion is mistaken. Instead, rates are more usefully thought of as parameters in models that predict the expected magnitude of change over time. Because different models imply different scalings of divergence with time, tempo is inextricably linked to the mode of evolution (Roopnarine 2003). Measuring rates of change therefore demands careful attention about how best to model morphological evolution on the time scales of interest.

It may be illustrative to consider the analogous problem of measuring rates of taxonomic origination and extinction. It was common, especially in early treatments, to measure extinction intensity by tabulating the number of taxa terminating in different intervals (e.g., Simpson 1944). A variety of metrics were devised subsequently to account for interval length, standing diversity, and other factors, but it was not clear which was optimal (Foote 2000; Alroy 2010). By contrast, current approaches for measuring taxonomic rates start with a simple branching model with per capita rates of origination and extinction and derive estimators appropriate to the nature of the data to be analyzed (Foote 2000, 2003; Alroy 2008; Liow and Nichols 2010). Thus, in modern paleontology, taxonomic rates are estimated model parameters, not ad hoc metrics.

Considering rates as model parameters has several important benefits. It allows one to

use standard statistical principles such as likelihood, with its advantages of good estimation performance, easy computation of confidence intervals, and well-established means for testing hypotheses and comparing models (Burnham and Anderson 2010; Wang 2010). In addition, employing explicit models facilitates comparisons across approaches, and across kinds of data. For example, taxonomic rates derived from the stratigraphic ranges of fossils can be compared with those derived from birth-death models applied to molecular phylogenies composed of strictly extant species (Nee 2004; Alfaro et al. 2009; Rabosky 2009).

These same advantages apply to the measurement of phenotypic rates of change. Whereas computing traditional rates in populations that do not form strict ancestor-descendant relationships is ambiguous (Hendry and Kinnison 1999), it is perfectly straightforward to estimate the random walk or stasis models in anagenetic lineages (Hunt 2006) and across species in a clade (Pagel 1998; Ackerly 2009; Harmon et al. 2010). The resulting estimates can be compared across paleontological and phylogenetic contexts, and it is straightforward to test whether rates of evolution are shared or different across traits, lineages or time periods (Hunt 2006, 2008b; O'Meara et al. 2006; Thomas et al. 2009; Geary et al. 2010).

### Practical Recommendations

*Use Likelihood, Not LRI.*—Simulations show that the LRI method does not estimate rates well under realistic paleontological conditions, even under simple models of trait evolution. The estimated generational rates are systematically biased by sampling error at low rates of change and they are always much less precise than those obtained via maximum likelihood. Sheets and Mitchell (2001b) and Roopnarine (2003) expressed doubts about whether LRI should be able to truly recover generational rates, but no systematic assessment has been published. Gingerich presented three realized sequences generated as random walks (one each in 1993, 2001, 2009), and in each case LRI recovered approximately the correct generating rate. However, none of these sequences were generated under condi-

tions identified here as contributing to difficulties in estimation. Two were analyzed at single-generation resolution, none included sampling error, and all involved high rates of change. Sequences were also quite long (50–200 samples), which should reduce estimation error. Simulations in the present study were performed under more realistic conditions of sequence length, sampling error, and temporal spacing of samples, and they were performed across rates ranging from nearly zero to quite high. For all simulated conditions, generational rates were estimated much more effectively by maximum likelihood than by LRI.

*Which Rate Metric Is Most Useful?*—The most sensible way to measure rate differs across these three models, and the diversity of evolutionary patterns among lineages suggests that no model—and therefore no rate metric—is universally appropriate. Which, then, is the best of the imperfect choices available for measuring rates?

Traditional rates (darwins, haldanes) do not explicitly assume a model, but they can be justified easily only when evolutionary change is strongly directional. Because directionality is rarely sustained over paleontologically relevant durations (Hunt 2007), traditional rate metrics are usually a poor choice for measuring the pace of phenotypic evolution. The rate metric considered here that is nearly equivalent to traditional rates,  $\mu^2_{sr}$ , has the strongest negative scaling with interval length. It may not be surprising that traditional rates, when used at all, are often limited to situations for which interval length can be controlled (e.g., [Smith et al. 2010](#)). When sequences vary more than slightly in resolution, much of the variation in traditional rate metrics will be accounted for by interval length, confounding attempts to compare the true pace of change.

Although traditional rate metrics are inadvisable, reasonable arguments can be made in support of using either the step variance or disparity approaches for measuring the pace of evolutionary change. Both follow from a model of evolution that is commonly supported in fossil lineages (random walk for  $\sigma^2_{sr}$ , stasis for  $\omega$ ), both have moderate

correlations with interval length, and both naturally account for evolutionary reversals and sampling error. The step variance has the additional benefits of widespread use in phylogenetic studies and a known range under neutral evolution.

Recent discussions of stasis emphasize that evolutionary differences between repeated samples do not cumulate into sustained divergence within species; sample-to-sample fluctuations are depicted as comparable in magnitude to the entire range of phenotypes explored through a lineage's history (Gould 2002; Eldredge et al. 2005). When this is an adequate description, divergence is best measured as disparity. However, such constancy of divergence seems to break down over shorter periods of time (Kinnison and Hendry 2001) and for evolution spanning clades rather than single lineages (Fig. 7C) (Uyeda et al. 2011).

I have focused here on a specific measure of disparity, the  $\omega$  parameter of the stasis model, but other measures of divergence are likely to have similar properties. One common alternative is the absolute magnitude of change,  $|\Delta X|$ . When standardized by the within-sample variance, this becomes equivalent to the numerator of the haldane rate metric. Relative to the haldane numerator,  $\omega$  has the advantages of incorporating information over all samples in a sequence and accounting for sampling error.

Although both  $\sigma_s^2$  and  $\omega$  have less severe temporal scaling than traditional rates (Table 2, Fig. 7), their correlations with interval length are strong enough to compromise comparisons across resolutions that span several orders of magnitude. Over narrower ranges of interval lengths, however, the effect is less severe. For example, consider within-lineage rates estimated over typical paleontological spans, from  $10^5$  to  $10^7$  generations ( $n = 167$ ). The rank correlation between  $\sigma_s^2$  and interval length for this subset is nearly zero ( $r_s = -0.015, p = 0.84$ ); for  $\omega$ , it is  $r_s = +0.269$  ( $p = 0.0004$ ). Thus, for rate comparisons in the current compilation that span typical paleontological resolutions, interval length has little influence on these rate metrics, especially for the step variance. By comparison, the rank correlation between  $\mu_s^2$

over this same limited range is much stronger ( $r_s = -0.429, p = 7 \times 10^{-9}$ ). When comparisons of interest vary widely in temporal span, it may be prudent to plot rates or divergences with respect to interval length to assess the degree to which the step variance and disparity metrics correlate with interval length. One can then choose the rate metric that is less time-dependent for a particular study.

It might seem safer to consider for a rate metric only those cases in which its underlying model is well supported. However, the temporal scaling of step variance in the subset of cases best interpreted as a random walk is similar to the scaling in the complete compilation, and the same is true looking at the disparity metric restricted to examples of stasis. Neither of these findings is expected if the generating models are exactly true, suggesting that even when a model is well supported, it is at best an approximation that may not apply to observations at different temporal scales.

Finally, I will close with two suggestions that may be worth exploring for comparing rates among sequences of different evolutionary modes. The first, suggested to me by Andrew Bush (personal communication 2011), would fit a model in which directional, random walk, and stasis components are all estimated, and to use these components as elements in a rate vector. Comparisons would then be made among sequences using the entire rate vector without attempting to reduce it to a scalar quantity. A second approach would take the best-supported model for each sequence separately and use the resulting parameter estimates to compute an expected ESD over a standard interval length (e.g., one million years). This predicted ESD could be compared across sequences regardless of their underlying mode of evolution. For example, if a sequence is best described as a random walk, then one takes the estimated step variance for that sequence and computes ESD from Table 1 as  $\sigma_s^2 t$ , substituting the standard interval length for  $t$ . Sequences best fit by the stasis model would do the parallel calculation using the expression  $2\omega$  for ESD, and so on. When the best fit model is also a good approximation at

the predicted interval length, this procedure is relatively safe; or, at least it is not clear how one would get a better prediction. A variant of this approach would compute the predicted ESD across all three models and then average them, weighted by the Akaike weight each model receives. Such an approach would be especially appropriate when no single model is unequivocally better than the others.

### Summary

1. Rates are most usefully viewed as parameters in models that relate the magnitude of divergence to elapsed time. The most appropriate way of measuring the rate of phenotypic change depends on the underlying mode of evolution. Random walks, strongly directional change, and stasis each imply different optimal strategies for measuring rates of evolution. When these models are true, their corresponding rate metric can be estimated fairly via maximum likelihood, regardless of the time scale over which they are measured.
2. Each of the rate metrics implied by these three models is equivalent to existing strategies in evolutionary biology. Assuming strongly directional evolution produces a metric closely related to traditional rate metrics such as darwins and haldanes. Random walks have a rate parameter called a step variance that determines the magnitude of change over a specified interval. The stasis model implies measuring "rate" without reference to time at all as the among-population variance in trait means, similar to the measure of disparity in morphospace studies.
3. The LRI method is an alternative approach that is said to estimate single-generation rates of change, regardless of the underlying mode of evolutionary change. Consideration of this approach, verified by extensive simulations, suggests that this method performs poorly in most realistic paleontological situations. Sampling noise causes this approach to overestimate rates unless sampling error

is very small. Over plausible ranges of true rates, LRI estimated rates are actually negatively correlated with true rates of change.

4. Analysis of a simple model with a population evolving near a peak in the adaptive landscape shows that generational rates cannot be inferred from observations made at much coarser resolutions. Such discontinuity between paleontological and generational patterns is probably expected under most realistic scenarios of evolutionary change.
5. Estimation of the three rate parameters across surveys of paleontological and phylogenetic data sets finds that the metric assuming strongly directional evolution has a strong negative scaling with interval length. The step variance and disparity approaches have weaker, but still significant, scalings with interval length. Their less severe temporal scaling makes the step variance and disparity metrics much more useful for measuring rates in empirical studies. The step variance metric has the additional benefits of a known range under neutral evolution and widespread use in phylogenetic comparative studies.

### Acknowledgments

I thank M. Foote and S. Wing for helpful discussions clarifying rates and their measurement; M. Foote also helped to improve an earlier version of this manuscript and suggested predicted ESD could be averaged across models. Thanks to L. Harmon for providing the raw data from his compilation of comparative phylogenetic data sets, and to H. D. Sheets and A. Bush for thoughtful and useful reviews of the present manuscript.

### Literature Cited

- Ackerly, D. 2009. Conservatism and diversification of plant functional traits: Evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences USA* 106:19699–19706.
- Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences USA* 106:13410–13414.

- Alroy, J. 2008. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences of the United States of America* 105:11536–11542.
- . 2010. Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. Pp. 55–80 in Alroy and Hunt 2010.
- Alroy, J., and G. Hunt, eds. 2010. *Quantitative methods in paleobiology*. Paleontological Society Papers 16. Paleontological Society, New Haven, Conn.
- Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macro-evolution. *Genetica* 112–113:9–32.
- Bookstein, F. L. 1987. Random walk and the existence of evolutionary rates. *Paleobiology* 13:446–464.
- Brusatte, S. L., S. Montanari, H. Y. Yi, and M. A. Norell. 2011. Phylogenetic corrections for morphological disparity analysis: new methodology and case studies. *Paleobiology* 37:1–22.
- Burnham, K. P., and D. R. Anderson. 2010. *Model selection and multimodel inference*. Springer, New York.
- Bush, A. 2000. Evolution within demes: stasis as a hierarchical phenomenon. *Geological Society of America, Abstracts with Programs* 32:445.
- Cheetham, A. H., J. B. C. Jackson, and L.-A. C. Hayek. 1993. Quantitative genetics of bryozoan phenotypic evolution. I. Rate tests for random change versus selection in differentiation of living species. *Evolution* 47:1526–1538.
- . 1994. Quantitative genetics of bryozoan phenotypic evolution. II. Analysis of selection and random change in fossil species using reconstructed genetic parameters. *Evolution* 48:360–375.
- Ciampaglio, C. N., M. Kemp, and D. W. McShea. 2001. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. *Paleobiology* 27:695–715.
- 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.
- Eldredge, N., and S. J. Gould. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Pp. 82–115 in T. J. M. Schopf, ed. *Models in paleobiology*. Freeman, Cooper, San Francisco.
- Eldredge, N., J. N. Thompson, P. M. Brakefield, S. Gavrilets, D. Jablonski, J. B. C. Jackson, R. E. Lenski, B. S. Lieberman, M. A. McPeck, and W. I. Miller. 2005. The dynamics of evolutionary stasis. *Paleobiology* 31(Suppl. to No. 2):133–145.
- Erwin, D. H., and R. L. Anstey. 1995. Speciation in the fossil record. Pp. 11–38 in D. H. Erwin and R. L. Anstey, eds. *New approaches to speciation in the fossil record*. Columbia University Press, New York.
- Erwin, D. H., M. Laflamme, S. M. Tweedt, E. A. Sperling, D. Pisani, and K. J. Peterson. 2011. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* 334:1091–1097.
- Estes, S., and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *American Naturalist* 169:227–244.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.
- Footo, M. 1991. Analysis of morphological data. Pp. 59–86 in N. L. Gilinsky and P. W. Signor, eds. *Analytical paleobiology*. Paleontological Society, Knoxville, Tennessee.
- . 2000. Origination and extinction components of taxonomic diversity: general problems. In D. H. Erwin and S. L. Wing, eds. *Deep time: Paleobiology's perspective*. *Paleobiology* 26(Suppl. to No. 4):74–102.
- . 2003. Origination and extinction through the Phanerozoic: a new approach. *Journal of Geology* 111:125–148.
- Geary, D. H., G. Hunt, I. Magyar, and H. Schreiber. 2010. The paradox of gradualism: phyletic evolution in two lineages of lymnocardiid bivalves (Lake Pannon, central Europe). *Paleobiology* 36:592–614.
- Gingerich, P. D. 1983. Rates of evolution: effects of time and temporal scaling. *Science* 222:159–161.
- . 1985. Species in the fossil record: concepts, trends, and transitions. *Paleobiology* 11:27–41.
- . 1993. Quantification and comparison of evolutionary rates. *American Journal of Science* 293-A:453–478.
- . 2001. Rates of evolution on the time scale of evolutionary process. *Genetica* 112–113:127–144.
- . 2009. Rates of evolution. *Annual Review of Ecology, Evolution, and Systematics* 40:657–675.
- Gould, S. J. 1984. Smooth curve of evolutionary rate: a psychological and mathematical artifact. *Science* 226:994–995.
- . 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.
- Haldane, J. B. S. 1949. Suggestions as to the quantitative measurement of rates of evolution. *Evolution* 3:51–56.
- Hannisdal, B. 2006. Phenotypic evolution in the fossil record: numerical experiments. *Journal of Geology* 114:133–153.
- Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351.
- Harmon, L. J., J. Weir, C. Brock, R. Glor, W. Challenger, and G. Hunt. 2009. *geiger: analysis of evolutionary diversification*, R package, Version 1.3–1.
- Harmon, L. J., J. B. Losos, T. J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. H. Kozak, M. A. McPeck, F. Moreno-Roark, T. J. Near, A. Purvis, R. E. Ricklefs, D. Schluter, J. A. Schulte, O. Seehausen, B. L. Sidlauskas, O. Torres-Carvajal, J. T. Weir, and A. O. Mooers. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385–2396.
- Hendry, A. P., and M. T. Kinnison. 1999. The pace of modern life: measuring rates of contemporary microevolution. *Evolution* 53:1637–1653.
- Hersch, E. I., and P. C. Phillips. 2004. Power and potential bias in field studies of natural selection. *Evolution* 58:479–485.
- Hunt, G. 2006. Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology* 32:578–601.
- . 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. *Proceedings of the National Academy of Sciences USA* 104:18404–18408.
- . 2008a. Evolutionary patterns within fossil lineages: model-based assessment of modes, rates, punctuations and process. In P. H. Kelley and R. K. Bambach, eds. *From evolution to geobiology: research questions driving paleontology at the start of a new century*. Paleontological Society Papers 14:117–131. Paleontological Society, Pittsburgh.
- . 2008b. Gradual or pulsed evolution: when should punctuational explanations be preferred? *Paleobiology* 34: 360–377.
- . 2011. *paleoTS: Modeling evolution in paleontological time-series*, R package, Version 0.4-1.
- Hunt, G., and M. T. Carrano. 2010. Models and methods for analyzing phenotypic evolution in lineages and clades. Pp. 245–269 in Alroy and Hunt 2010.
- Jablonski, D., and D. J. Bottjer. 1990. Onshore-offshore trends in marine invertebrate evolution. Pp. 21–75 in W. D. Allmon and R. M. Ross, eds. *Causes of evolution: a paleontological perspective*. University of Chicago Press, Chicago.
- Kendall, M. G., and A. Stuart. 1969. *The advanced theory of statistics, Vol. 1. Distribution theory*. Hafner, New York.

- Kingsolver, J. G., and D. W. Pfennig. 2007. Patterns and power of phenotypic selection in nature. *BioScience* 57:561–572.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gilbert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. *American Naturalist* 157:245–261.
- 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.
- Knapczyk, F. N., and J. K. Conner. 2007. Estimates of the average strength of natural selection are not inflated by sampling error or publication bias. *American Naturalist* 170:501–508.
- Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30:314–334.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Liow, L. H., and J. D. Nichols. 2010. Estimating rates and probabilities of origination and extinction using taxonomic occurrence data: capture-mark-recapture (CMR) approaches. Pp. 81–94 in Alroy and Hunt 2010.
- Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *American Naturalist* 136:727–741.
- Marshall, C. R. 2006. Explaining the Cambrian “explosion” of animals. *Annual Review of Earth and Planetary Sciences* 34:355–384.
- Mousseau, T. A., and D. A. Roff. 1987. Natural selection and the heritability of fitness components. *Heredity* 59:181–197.
- Nee, S. 2004. Extinct meets extant: simple models in paleontology and molecular phylogenetics. *Paleobiology* 30:172–178.
- O’Meara, B. C., C. c. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922–933.
- Pagel, M. 1998. Inferring evolutionary processes from phylogenies. *Zoologica Scripta* 26:331–348.
- Polly, P. D. 2002. Phylogenetic tests for differences in shape and the importance of divergence times: Eldredge’s enigma explored. Pp. 220–246 in N. MacLeod and P. L. Forey, eds. *Morphology, shape and phylogeny*. Taylor and Francis, London.
- Rabosky, D. L. 2009. Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. *Systematic Biology* 58:629–640.
- Roopnarine, P. D. 2003. Analysis of rates of morphologic evolution. *Annual Review of Ecology and Systematics* 34:605–632.
- Roy, K., and M. Foote. 1997. Morphological approaches to measuring biodiversity. *Trends in Ecology and Evolution* 12:277–281.
- 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.
- Simpson, G. G. 1944. *Tempo and mode in evolution*. Columbia University Press, New York.
- Smith, F. A., S. M. Elliott, and S. K. Lyons. 2010. Methane emissions from extinct megafauna. *Nature Geoscience* 3:374–375.
- Stanley, S. M. 1985. Rates of evolution. *Paleobiology* 11:13–26.
- Thomas, G. H., S. Meiri, and A. B. Phillimore. 2009. Body size diversification in *Anolis*: novel environment and island effects. *Evolution* 63:2017–2030.
- Uyeda, J. C., T. F. Hansen, S. J. Arnold, and J. Pienaar. 2011. The million-year wait for macroevolutionary bursts. *Proceedings of the National Academy of Sciences USA* 108:15908–15913.
- Wang, S. C. 2010. Principles of statistical inference: likelihood and the Bayesian paradigm. Pp. 1–18 in Alroy and Hunt 2010.

## Appendix 1

## Expected Generational Rate from Sampling Error

We wish to determine the expected generational rate in haldanes in the absence of any real evolutionary change. The rate in haldanes over a single generation is equal to  $|X_2 - X_1|/\sqrt{V_P}$  where the  $X_1$  and  $X_2$  are the observed trait means in generations 1 and 2, and  $V_P$  is the variance of the trait in each population (assumed to be the same across populations). Even though the two generations have the same true mean, with finite samples of  $n$  individuals, the estimated means,  $X_1$  and  $X_2$ , will not be exactly the same. Sampling error for any estimated mean is normally distributed with a mean of zero and a variance equal to  $V_P/n$ . The sampling variance of the difference between two means is the sum of their separate sampling variances, and so  $X_2 - X_1$  will be Gaussian with a mean of zero and variance of  $2V_P/n$  (Hunt 2006). Dividing by the haldane denominator  $\sqrt{V_P}$  means that  $(X_2 - X_1)/\sqrt{V_P}$  will have a variance of simply  $2/n$ . The expected absolute value of this rate is equal to the mean absolute deviation (MAD) of this scaled difference. The MAD of a Gaussian variable is equal to  $\sqrt{(2/\pi)}$  times its standard deviation, and thus the expected generational rate in the absence of evolution is

$$\sqrt{2/\pi}\sqrt{2/n} = 2/\sqrt{\pi n}.$$

Biometric studies of fossil lineages typically have average sample sizes that range from about  $n = 10$  to  $n = 50$ , which translates to generational rates from 0.16 to 0.36. This range is quite close to the typical range reported for generational rates (Polly 2002; Gingerich 2009). For example, the position of the entoconid cusp in *Cantius* (Fig. 3) shows little evolutionary change and has an average sample size of 7.65. From the equation above, this should produce a generational rate of 0.41, which is very close to the observed rate of 0.43.

These calculations predict the expected rate in haldanes observed from one generation to the next. For paleontological data, the single-generation rates obtain from extrapolating the log-rate, log-interval regression and this extrapolation has additional effects as described in the text.

## Appendix 2

## Expected Squared Divergence for an OU Processes

For an Ornstein-Uhlenbeck process, Hansen (1997) gives the expected (mean) trait value over time:

$$M(X) = [1 - \exp(-\alpha t)]\theta + \exp(-\alpha t)X_0,$$

with parameters as defined in the text. The expected change in a trait,  $\Delta X$ , is equal to the above expression minus the initial trait value,  $X_0$ , which simplifies to

$$M(\Delta X) = [1 - \exp(-\alpha t)](\theta - X_0). \quad (1)$$

Again from Hansen (1997), the variance of the trait value, which is also the variance of the change in trait values from any specified starting point, is

$$V(\Delta X) = \frac{\sigma_s^2}{2\alpha} (1 - \exp[-2\alpha t]). \quad (2)$$

The expected squared divergence (ESD) is equal to  $M^2 + V$ , substituting in the specific expressions from the previous two equations for  $M$  and  $V$ . What results is a complex expression of all four parameters that does not readily simplify.

The text explores a more specific scenario in which the initial trait value is equal to the optimum and  $\theta - X_0$  is therefore equal to zero. Under this condition,  $M(\Delta X) = 0$  and ESD simplifies to equation (2). Figure 6 shows plots of this expression as a function of time for several combinations of  $\alpha$  and  $\sigma_s^2$ . The behavior of this function can be understood as follows. As long as selection is not

too strong, then over very short time scales (small  $\alpha t$ ),  $\exp(-2\alpha t) \approx 1 - 2\alpha t$ , and the ESD from equation (2) is approximately equal to  $\sigma_s^2 t$ . Thus, initial divergence is proportional to the input from genetic drift as measured by the step variance. After a long time has elapsed ( $t$  very high), the part of equation (2) inside the

parentheses converges to one, and divergence asymptotes to a value equal to  $\sigma_s^2 / (2\alpha)$ . Divergence at very short time scales approximately follows a random walk, gradually transitioning over coarser and coarser resolutions to stasis-like behavior (with  $\omega = \sigma_s^2 / [2\alpha]$ ).