Genetica **112–113:** 127–144, 2001.
© 2001 *Kluwer Academic Publishers. Printed in the Netherlands.*

Rates of evolution on the time scale of the evolutionary process

Philip D. Gingerich
Departments of Geological Sciences, Biology, Anthropology, Museum of Paleontology, The University of Michigan, Ann Arbor, MI 49109-1079, USA

Key words: rates of evolution, temporal scaling, generational time scale, microevolution, macroevolution

Abstract

A generational time scale, involving change from one generation to the next, is the time scale of evolution by natural selection. Microevolutionary and macroevolutionary patterns reflect this process on longer time scales. Rates of evolution are most efficiently expressed in haldane units, H, in standard deviations per generation, indexed by the log of the time interval. Rates from replicated selection experiments and simulations have rate-interval [RI] and log rate-log interval [LRI] scaling relations enabling directional, stationary, and random time series to be distinguished. Empirical microevolutionary and macroevolutionary data exhibit stationary scaling, but point to generational rates of evolution (H_0) conservatively on the order of 0.2 standard deviations per generation on the time scale of the evolutionary process. This paradox of long-term stationary scaling and short-term high rates of change can be explained by considering the shape of an heuristic time-form evolutionary lattice. Cenozoic mammals occupy a lattice that is about four orders of magnitude longer in time than it has ever been wide in form. The evolutionary process is dynamic but operates within relatively narrow morphological constraints compared to the time available for change.

Introduction

Evolution is the name we give both to *patterns of change* in the forms of life we observe, and to the *process of natural selection* that produces these patterns. Dobzhansky (1937, p. 12) distinguished *microevolution*, representing evolution observable in the span of a human lifetime, from *macroevolution*, involving patterns on a geological time scale. Generally speaking, the former includes experimental and historical patterns that may span anywhere up to 100 or so generations, while the history of life over geological time involves patterns on scales of hundreds to millions of generations. However, there are really three time scales of interest in evolution:

1. a *generational* time scale, involving change from generation-to-generation;
2. a *microevolutionary* time scale, involving patterns on scales from 2–100 or so generations; and
3. a *macroevolutionary* time scale, involving patterns on scales of hundreds or more generations.

The first of these is the time scale of the evolutionary process, and the second and third show how the process plays out in producing patterns of change over longer and longer time scales. There are relatively few studies quantifying microevolutionary change, and many more quantifying macroevolutionary change. The former sometimes involve artificial selection, while the latter usually treat organisms that lived before humans and are clearly free of human influence. The former yield higher rates of change and the latter yield lower rates: which studies – microevolutionary or macroevolutionary (or neither) – yield rates representative of the evolutionary process on a generational scale of time? We cannot understand the evolutionary process without answering this fundamental question.

Quantification of rates

Charles Darwin's *Origin of Species* (1859, p. 116–117) included a single figure, showing a hypothetical

pattern of what he considered 'species... supposed to resemble each other in unequal degrees' and their 'varying offspring' changing through time, where the intervals of time 'may represent each a thousand generations'. The degrees of resemblance and variation were not explicitly quantified, but there is a sense in Darwin's diagram of resemblance-standardized-by-variation, and the time scale is in generations. Ninety years later Haldane proposed (1949, p. 55) that 'it may be desirable to coin some word, for example a *darwin*, for a unit of evolutionary rate, such as an increase or decrease of size by a factor of e per million years', where e is the base of the natural logarithms. Haldane then found that horses evolved at rates of about 40 millidarwins (0.40 darwins), calculated on time scales of 5–14 million years. Rates have generally been calculated in darwins and millidarwins ever since.

However, Haldane (1949, p. 52) prefaced his proposal of the darwin unit by noting that 'if evolution... depends on selection, the generation seems the more suitable unit of time', and 'the use of the standard deviation as a yardstick has a certain interest because, on any version of the Darwinian theory, the variation within a population at any time constitutes, so to say, the raw material available for evolution'. Following Haldane, Lerman (1965), Lande (1976), Lynch (1990), and others, it is desirable to quantify evolutionary change in terms of proportion expressed in units of standard deviation per generation. I have called this unit a *haldane* (Gingerich, 1993), both to recognize it as Haldane's and to contrast it with his widely used darwin.

Calculating a rate of change between two samples requires three quantities: (1) the proportional difference between the sample means, $d = \bar{y}_2 - \bar{y}_1$; (2) the pooled standard deviation of the samples, e.g., $s_\mathrm{p} = \sqrt{s_\mathrm{p}^2}$, where $s_\mathrm{p}^2 = ((n_1-1)s_1^2+(n_2-1)s_2^2)/(n_1+n_2-2)$ (Sokal & Rohlf, 1981, p. 226); and (3) the time interval between the samples, $I = t_2 - t_1$, counted or estimated in generations. Proportionality is incorporated by taking $\bar{y}_\mathrm{i}$ to be the mean of each sample of logged measurements, while s_i is the standard deviation of each sample of logged measurements (logs to the base e, ln, are generally used because they make the standard deviation of the logged measurements equivalent to the more familiar coefficient of variation of unlogged measurements; Lewontin, 1966). Logging is necessary because of the geometric normality of biological variation (Gingerich, 2000). Finally, the rate of change in standard deviations per generation (haldanes) is calculated as:

$$H_{(\log I)} = D/I, \quad \text{where} \quad D = d/s_\mathrm{p}. \tag{1}$$

Rates calculated over different intervals are not directly comparable, which is why rates must be indexed by the log (here base 10) of the interval.

Rates calculated in haldanes and rates calculated in darwins are correlated (Hendry & Kinnison, 1999), but they are not the same. Rates in darwins, while dimensionless numbers, retain the same proportionality to dimension as observed for coefficients of variation (which can be confirmed by simple calculations paralleling those of Lande, 1977). Rates in haldanes are independent of dimension. Quantification in haldanes is always more appropriate in evolutionary studies because, as Haldane himself stated, variation is the raw material of evolution and a generational time scale is the time scale on which evolution takes place.

Variation is essential for evolution of any kind, whether this be by artificial selection, natural selection, or random drift, and this variation must be quantified so rates can be appropriately scaled. Change by any absolute amount or by any proportion is only meaningful in terms of the distribution of variation available for selection or drift: this is the basis for the concept of selection intensity (see below). The standard deviation, the root-mean-square of all deviations and the distance from the mean of a normal curve to either of its inflection points, is a natural measure of variation whatever the original unit of measurement.

A generational time scale is the time scale on which the process of evolution takes place. The evolutionary process has no memory of a past beyond that represented in genes and morphology of the present generation, nor any clairvoyant anticipation of a future: any selective decisions about what serves an organism advantageously and disadvantageously are made here-and-now as it interacts with its environment from conception through growth to reproduction. Hence in characterizing the process there is no alternative to knowing or inferring rates on a generation-to-generation time scale.

Evolutionary process: a model for selection

A graphic model for quantification of natural selection is shown in Figure 1. This was developed as a teaching tool, substituting a simple linear selection gradient for the usual textbook case of selection by

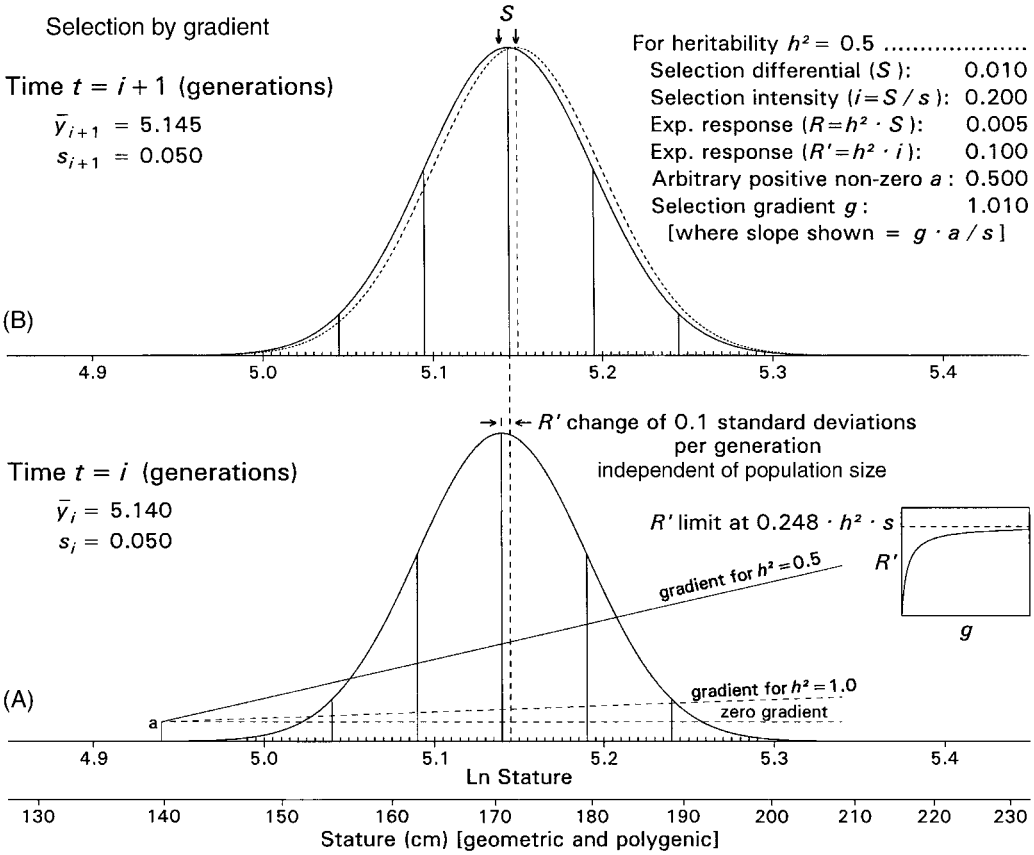

Figure 1. Change in human stature as a model for evolution from one generation (A) to the next (B) in response to a gradient of selection (represented by the diagonal line superimposed in A). Population samples are normal on a logarithmic scale, with means of 5.140 and 5.145 ln units (corresponding to statures of 170.7 and 171.6 cm, respectively) and a common standard deviation of 0.05 ln units. Vertical lines within normal curves are standard deviation [s.d.] units. Heritability h^2 is assumed to be 0.5, which means that a selection differential S of 0.01 ln units, equivalent to a selection intensity i of 0.2 standard deviation units (dashed normal curve in B), is required to achieve a response of 0.005 ln units (R) – equivalent to the response of 0.1 s.d. units shown here (R'; solid normal curve in B). Selection gradient in A indicates how sizes are favored in relation to an arbitrarily small positive non-zero constant a at $\bar{x} - 4$ s.d. The gradient g shown here is that required to achieve a change in means of 0.1 s.d. units for $h^2 = 0.5$ (compare to the corresponding gradient for $h^2 = 1.0$). This change is *independent of population size* and equally efficient for populations of small or large effective size. The expected response R' has a limit of approximately 0.124 standard deviation units in this example (inset box at right in lower panel shows how R' is related to g). Documentation of higher rates indicates that such a linear gradient is probably not representative and truncation selection may be more common in nature than is generally recognized. The selection and response notation here follows Falconer (1981; with a added to scale gradient slope to unit area under normal curve).

truncation (e.g., Falconer, 1981, p. 174; Roff, 1997, p. 119). It is the simplest linear univariate reduction of Lande and Arnold's (1983) more general multivariate treatment. The model illustrates selection in the context of the well-documented difference in stature of my generation (Figure 1(A)) and that of my students (Figure 1(B)) – the latter being, on average, about a centimeter taller. In the model, my generation at time i is represented as having a mean stature $\bar{y}_i = 5.140$ on a natural logarithmic scale (equivalent to about 171 cm on a centimeter scale of measurement), with a standard deviation $s_i = 0.50$. My students' generation at time $i + 1$ is represented as having a mean stature $\bar{y}_{i+1} = 5.145$ on a natural logarithmic scale (equivalent to about 172 cm on a centimeter scale of measurement), with a standard deviation $s_{i+1} = 0.50$.

The difference in stature in successive generations can be explained by imagining a linear gradient of selection affecting my generation, giving some slight statistical reproductive advantage to taller people. The slope of the gradient here is given by a/s, where a is an arbitrarily small constant and s is a standard deviation unit (a is required to scale the slope relative to unit area under the normal curve). A zero gradient (lower dashed line) would yield no effect, while a slightly steeper gradient (upper dashed line) would

be sufficient to explain the observed shift of means if heritability h^2 were complete ($h^2 = 1.0$). The gradient required for a more realistic case of partial heritability ($h^2 = 0.5$) is shown with a solid line.

A selection differential S of 0.01 ln units, equivalent to a selection intensity i of 0.2 standard deviation units, is required to effect an expected response R = 0.005 in ln units and $R' = 0.1$ in standard deviation units. Interestingly, in the model this response has a limit of $0.248 \cdot h^2 \cdot s$ (inset graph in Figure 1(A)) for a linear selection gradient, which is only 0.124 standard deviations per generation when $h^2 = 0.5$. Thus it is possible to explain an increase of a centimeter or so of human stature in a single generation with a gradient of selection, but the resulting rate $H_0 \approx 0.1$ is close to the upper limit expected for a linear selection gradient.

This example illustrates: (1) how the process of evolution by natural selection takes place on a generation-to-generation time scale; and (2) how natural it is to quantify change in standard deviation units per generation, whether the quantity of interest is expected response R' or calculated rate of change H_0. Anthropologists call the documented change leading to greater human stature in recent generations the 'human secular trend' but claim that it cannot represent evolution 'because it is too fast'. I do not know if this secular trend is evolutionary (nor do I claim this – its heritability has not been demonstrated and the trend could be due simply to improved nutrition), but as I shall show below, such change cannot be ruled out as evolution because it is too fast.

Evolutionary rates: experimental selection in the laboratory

Evolution as a process takes place on a generational time scale, repeated generation after generation until patterns of change are produced on longer scales of microevolutionary and macroevolutionary time. The transition from generational to longer time scales is amenable to investigation through experimentation and simulation, and in rare instances field observation. Falconer's (1973) replicated selection experiments for body weight in mice provide an example of experimental evolution (Figure 2(A)). Three experiments were carried out in parallel. In the first, six 'large' lineages were maintained in the laboratory for 23 generations, selected in each generation for larger body size. In the second, six 'small' lineages were maintained in the laboratory for 23 generations, selected in each generation for smaller body size. In the third, six 'control' lineages were maintained in the laboratory for 23 generations, with founders of the next generation drawn at random.

Each of Falconer's lineages yields 23 independent rates of response on a 1-generation time scale, and each set of six lineages yields 138 independent rates on a 1-generation time scale (Figure 2(A)). These are *autonomous rates* because in each lineage the time intervals are independent and non-overlapping. It is possible to calculate rates on longer 2-generation, 3-generation, etc., time scales by comparing intervals from t to t_2, t_1 to t_3, etc., but these would not be autonomous rates because time intervals overlap others in the same lineage and hence are not independent of each other. An example of nonautonomous rates spanning ten generations is shown in Figure 2(B), and an example of rates spanning all 23 generations is shown in Figure 2(C). The latter could be considered autonomous by themselves, but they are not independent of the rates shown in Figure 2(A).

The rates that can be calculated for each of Falconer's experiments are shown graphically in Figures 3–5. The 'large' experiment has 138 1-generation rates with a mean rate on this scale of 0.111 and a standard deviation on this scale of 0.385 (Figure 3(A)), 132 2-generation rates with a mean of 0.110 and standard deviation of 0.212 (not shown), 126 3-generation rates with a mean of 0.111 (Figure 3(B)), etc. Note that the mean is relatively stable, while the standard deviation decreases rapidly at first and continues decreasing as interval size gets longer. This is easily seen in the histograms, and also in the rate versus interval [RI] graph in Figure 3(M). A line fit to all points declines slightly as interval length gets longer (slope −0.001), indicating a very slight decline in response over time. This line has an intercept of 0.109, and an expected H_0 of 0.108 (value of regression for an interval of one generation), which is very close to the observed mean rate of 0.111 for 1-generation rates (Figure 3(A)). A different rate fingerprint is shown in the log rate versus log interval [LRI] graph in Figure 3(N). Logging accentuates the slight negative slope of the line in the previous panel, and this has a slope of −0.236 on the LRI graph. The intercept of an ordinary LRI plot is not a good predictor of H_0 because averaging positive and negative rates separately or pooling them as absolute values systematically overestimates the median.

The 'small' experiment (Figure 4) has 138 1-generation rates with a mean rate on this scale of

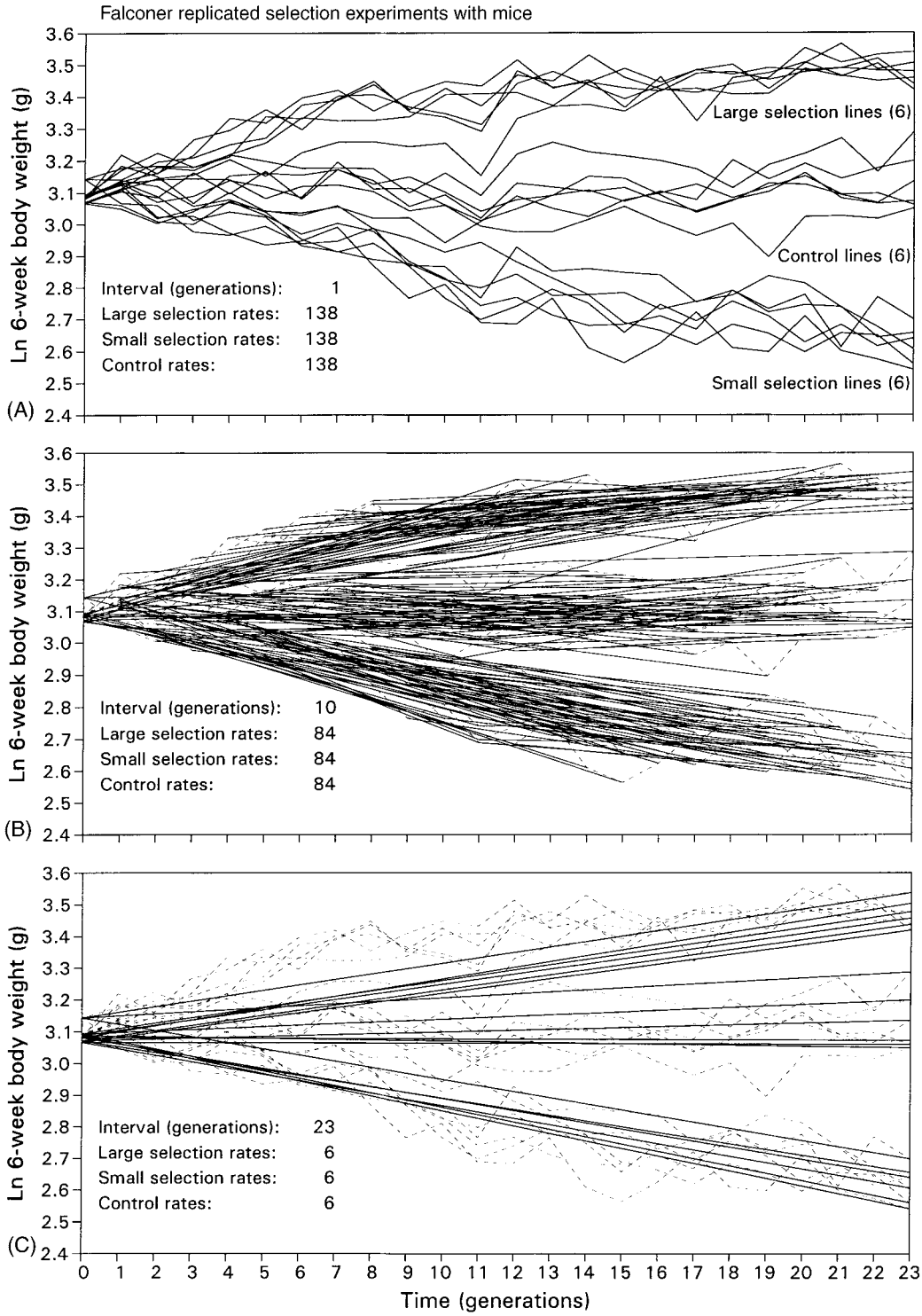

Figure 2. Replicated selection experiments of D.S. Falconer (1973) involving 6-week body weight (g) of laboratory mice. Six inbred populations were divided, with each subgroup founding a line selected in each generation to be large, selected to be small, or bred without artificial selection to serve as a control. Experiments were continued for 23 generations. (A) Generation means for each of the six lines in each experiment (large, small, and control, respectively): rates of response are indicated by the slope of each line segment. (B) Rates of response calculated at 10-generation intervals. (C) Rates of response calculated at 23-generation intervals. Rates of response on all time scales are compared in Figures 3–6.

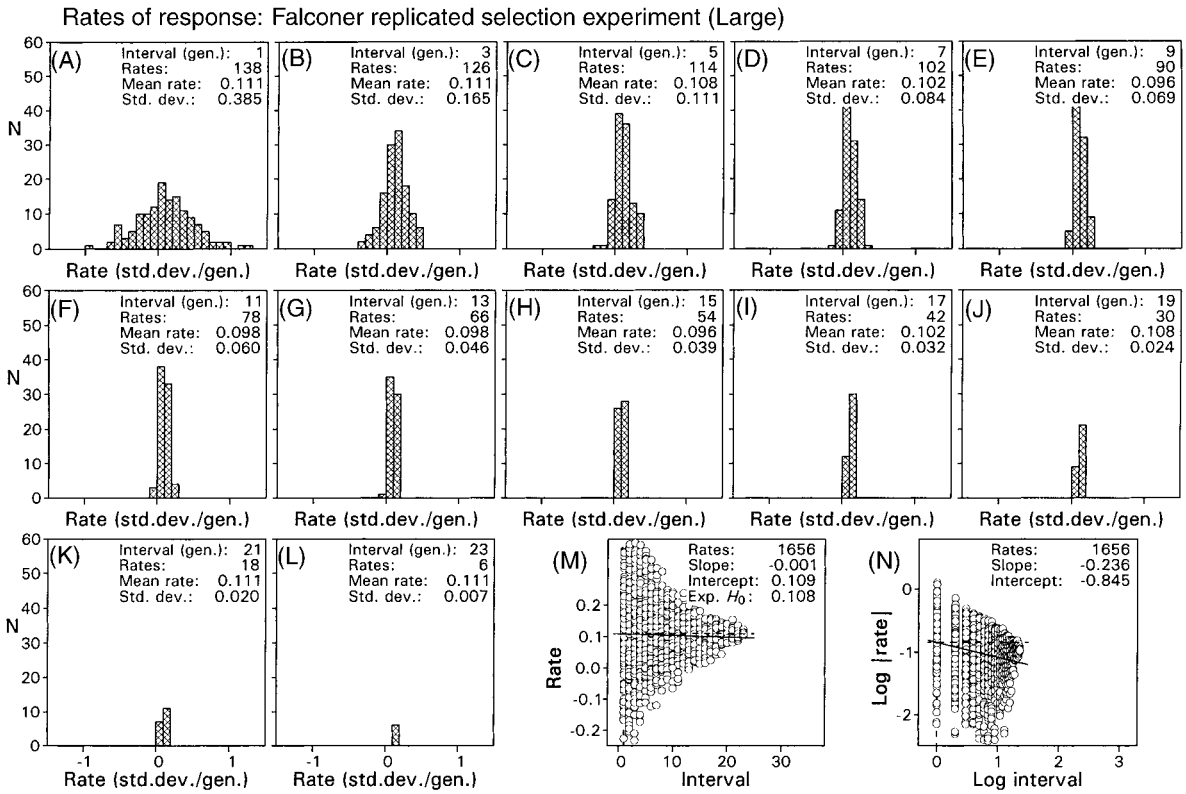

Figure 3. Comparison of rates of response in the Falconer experiment involving generational-scale selection for large body size. (A–L) Histograms of rates calculated over different intervals of time (panels for even-numbered intervals are omitted to conserve space) – note the symmetry of mean rates in panels A and L, B and K, etc., while standard deviations decline systematically. (M) Rate versus interval [RI] plot combining rates for all time scales – predicted H_0 is value of regression for an interval of one generation. (N) Log absolute value of rate versus log interval [LRI] plot. Note that rates in LRI plot decline slightly with interval length (solid line slope = −0.236) compared to expectation for a purely directional process (dashed line slope = 0.000).

−0.137 and a standard deviation on this scale of 0.404 (Figure 4(A)), 132 2-generation rates with a mean of −0.140 and standard deviation of 0.227 (not shown), 126 3-generation rates with a mean of −0.138 and standard deviation of 0.160 (Figure 4(B)), etc. Note that again the mean is relatively stable, while the standard deviation decreases rapidly at first and continues decreasing as interval size gets longer. This is easily seen in the histograms, and also in the rate versus interval [RI] graph in Figure 4(M). A line fit to all points declines slightly as interval length gets longer (slope −0.000), indicating a very slight decline in response over time. This line has an intercept of −0.138, and a predicted H_0 of −0.138, which is again very close to the observed mean rate of −0.137 for 1-generation rates (Figure 4(A)). Again, logging accentuates the slight negative slope of the line in the previous panel, and this has a slope of −0.145 on the LRI graph (Figure 4(N)).

The 'control' experiment (Figure 5) has rates distributed similarly to those in the two selection experiments. A line fit to all points in the RI graph (Figure 5(M)) declines slightly as interval length gets longer (slope −0.000), indicating a very slight decline in response over time. This line has an intercept of 0.004, and a predicted H_0 of 0.003, which is again close to the observed mean rate of 0.012 for 1-generation rates (Figure 5(A)). However, logging yields a much more negative slope representing the relationship of log rate to log interval, and this has a slope of −0.837 on the LRI graph (Figure 5(N)).

For completeness I constructed a fourth replicated selection experiment by simulation, with the response in each generation determined by selection at random from a normal distribution of rates having the parameters of 1-generation rates in the control experiment (Figure 5(A)). The 'random' experiment (Figure 6) has rates distributed like those in the three previous

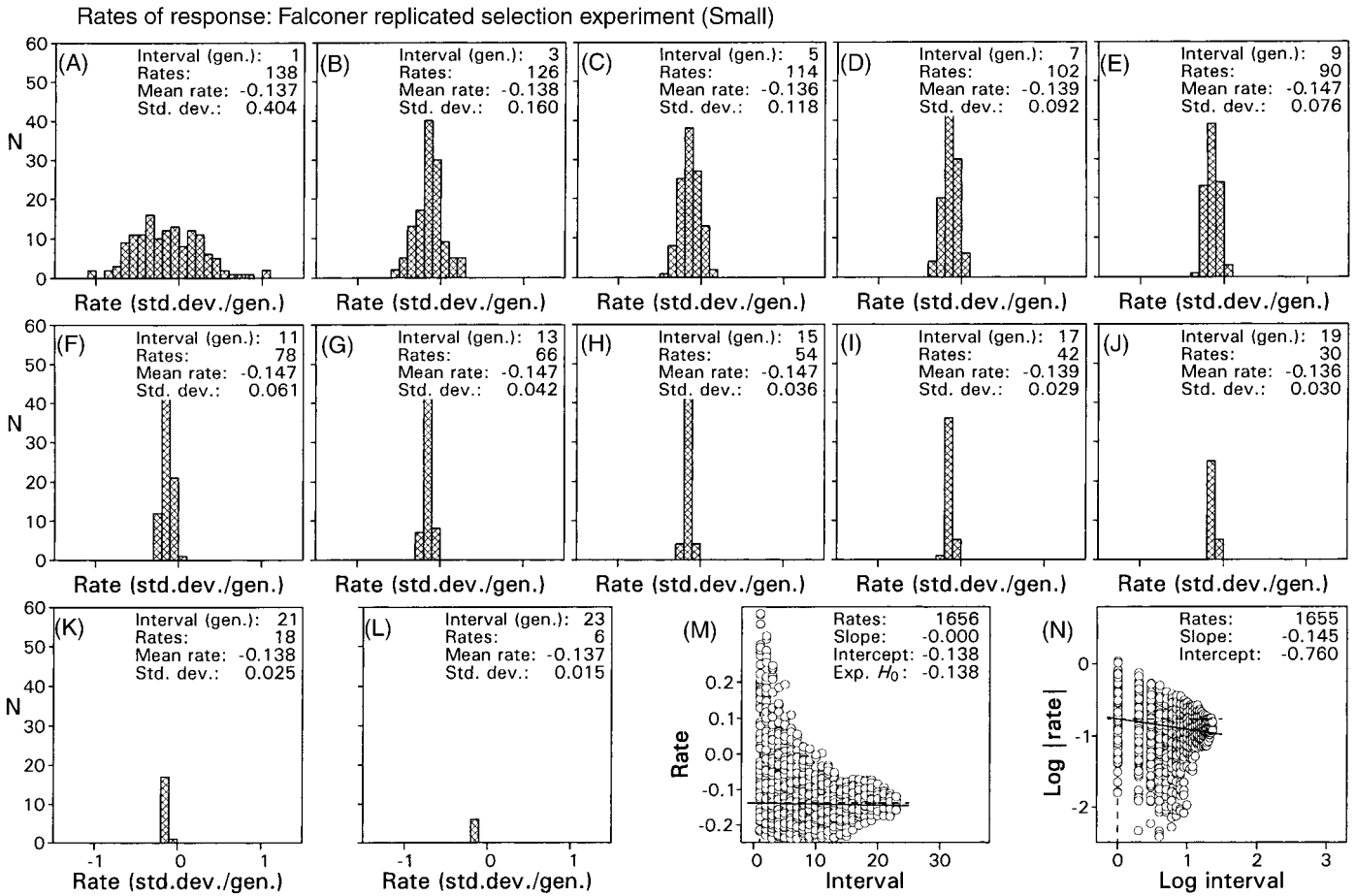

Figure 4. Comparison of rates of response in the Falconer experiment involving generational-scale selection for small body size. (A–L) Histograms of rates calculated over different intervals of time (panels for even-numbered intervals are omitted to conserve space) – note the symmetry of mean rates in panels A and L, B and K, etc., while standard deviations decline systematically. (M) Rate versus interval [RI] plot combining rates for all time scales – predicted H_0 is value of regression for an interval of one generation. (N) Log absolute value of rate versus log interval [LRI] plot. Note that rates decline slightly with interval length (solid line slope = −0.145) compared to expectation for a purely directional process (dashed line slope = 0.000).

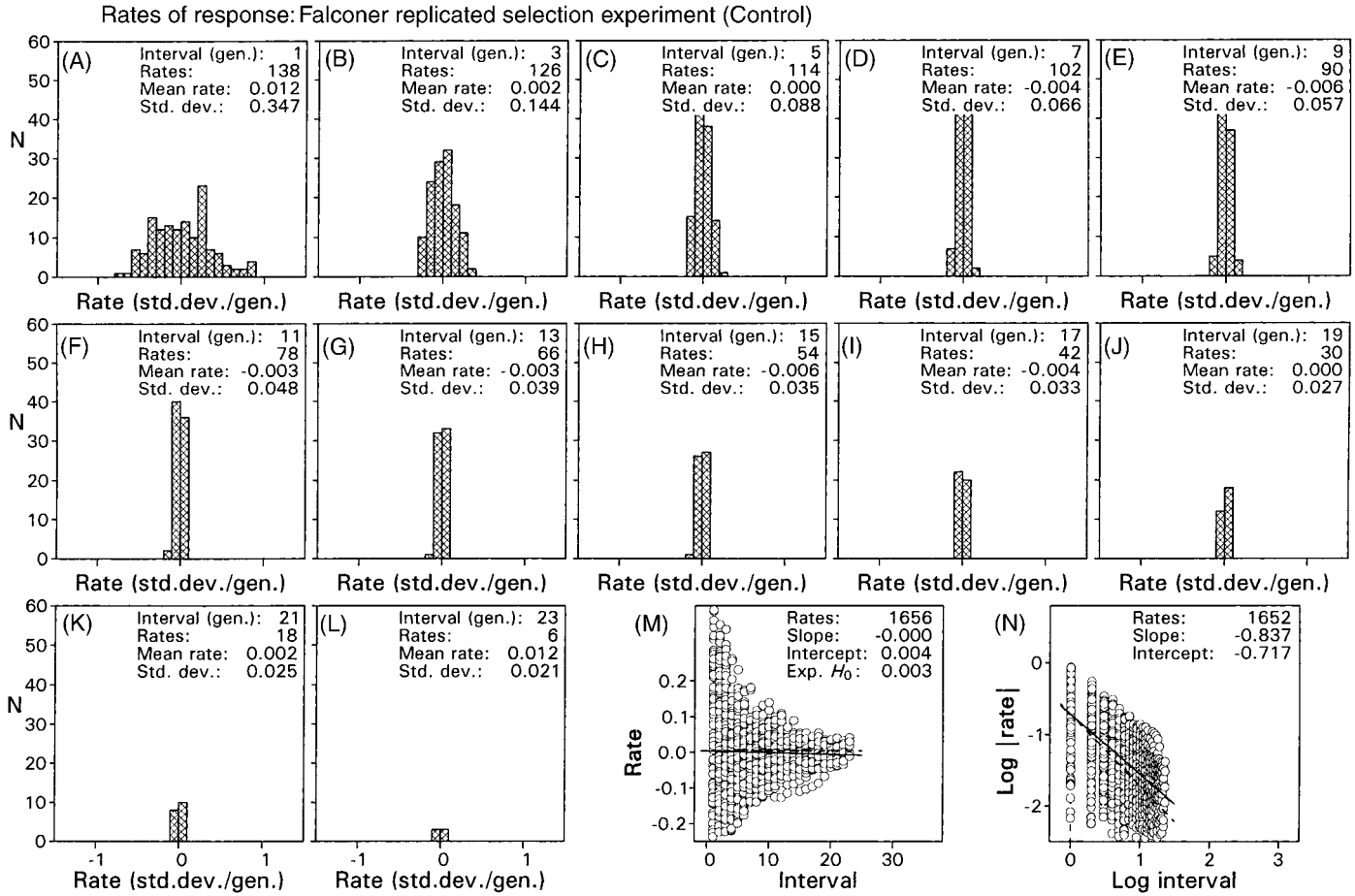

Figure 5. Comparison of rates of response in the Falconer control experiment involving no artificial selection. (A–L) Histograms of rates calculated over different intervals of time (panels for even-numbered intervals are omitted to conserve space) – note the symmetry of mean rates in panels A and L, B and K, etc., while standard deviations decline systematically. (M) Rate versus interval [RI] plot combining rates for all time scales – predicted H_0 is value of regression for an interval of one generation. (N) Log absolute value of rate versus log interval [LRI] plot. Note that rates decline steeply with interval length (solid line slope $= -0.837$) but exceed expectation for a purely stationary process (dashed line slope $= -1.000$).

experiments. A line fit to all points in the RI graph (Figure 6(M)) inclines slightly as interval length gets longer (slope 0.001), indicating a very slight chance increase in response over time. This line has an intercept of 0.036, and a predicted H_0 of 0.037, which is close to the observed mean rate of 0.030 for the simulated sample of 1-generation rates. Again logging yields a negative slope representing the relationship of log rate to log interval, and this has a slope of −0.550 on the LRI graph (Figure 6(N)).

Evolutionary rates: three time-series simulations

The easiest way to understand the rate-interval fingerprints in panels M and N of Figures 3–6 is through analysis of response rates in time series constructed by simulation to satisfy clearly-specified constraints. Three time series are shown in Figure 7(A). Each of these has a generation rate (base rate or step rate) set arbitrarily at a constant value of ±0.1 standard deviation units per generation (this value need not be constant, and it cannot be both constant and zero). The first time series is a directional trend constrained to have the same positive sign at every step (solid squares in Figure 7(A)). The second time series is stationary, with the sign free to change but constrained to change at every step (solid circles in Figure 7(A)).

The third time series is random, with the sign free to change at every step and the sign determined by chance independently of the sign of the previous step (open circles in Figure 7(A)).

Each time series in Figure 7(A) has distinct RI and LRI fingerprints shown in Figures 7(B)–(C). The directional trend (solid squares) is so constrained that its rate $r = 0.1$ is independent of the interval $(t_i - t_{i-j})$ over which it is calculated, and its RI fingerprint is just a straight line of zero slope with intercept 0.1. Logging does not affect the independence and the LRI fingerprint of this directional trend is also just a straight line of zero slope with intercept $10^{-1} = 0.1$.

The stationary time series (solid circles) has a rate r that depends on the interval $(t_i - t_{i-j})$ so completely that r is just the scaled inverse of the interval, and its RI fingerprint (Figure 7(B)) is an inverse power distribution centered on a line of zero slope with intercept close to zero (here 0.002). The LRI fingerprint of the stationary time series (Figure 7(C)) is a straight line of slope −1.0 and intercept $10^{-1} = 0.1$.

The random time series (open circles) has a rate r that depends on the square root of the interval $(t_i - t_{i-j})$. Its RI fingerprint (Figure 7(B)) is again an inverse power distribution centered on a line of zero slope. However the RI intercept of this center line is −0.025, reflecting the negative drift of the particular random series shown. The LRI fingerprint of the random time series is a scatter of points with slope −0.360 and intercept $10^{-1} = 0.1$. The expected slope for a random time series is −0.500 because statistically y is proportional to $\sqrt{t}$, and the rate y/t is thus proportional to $t^{-0.5}$.

All three RI fingerprints can be modeled as power functions of the general form:

$$R = \pm r \cdot I^{-m}, \tag{2}$$

where R is the rate ordinate, r is the generational or step rate of the time series, I is the interval abscissa, and $-m$ is a variable in the range $-1 \leq -m \leq 0$. Variable $-m = 0$ for the directional time series, $-m = -1$ for the stationary time series, and $-m \approx -0.5$ for the random time series.

The three LRI graphs in Figure 7(C) can be derived from the RI graphs in Figure 7(B). After logging, equation 2 becomes:

$$\log R = \log \mid \pm r \mid - m \cdot \log I, \tag{3}$$

which is a linear equation in log I and log R, with slope $-m$ and intercept log r. Parameter r is 0.1 and log $r = -1$. Slope m has expected values of 0, −1, and −0.5 as before.

Another way to look at LRI slopes is as indicators of fractional or fractal dimension (Mandelbrot, 1967), where fractal dimension $D = |-m| + 1$. A directional time series has $D = 1$, the dimension of an ordinary Euclidean line. A stationary time series has $D = 2$, the dimension of an ordinary Euclidean area, reflecting how a zig-zag stationary time series effectively fills a two-dimensional area. Finally, a random time series has $D = 1.5$, which is the intermediate fractal dimension associated with a simple random walk (Mandelbrot, 1983, p. 240). The idea that rates of change are often inversely related to interval length is well established in other contexts (e.g., Sadler, 1981, 1993; McShea & Raup, 1986; Foote, 1994).

The most effective way to characterize rates of change on a generational time scale, H_0, for a sample rate distribution like those in Figure 7(B) is to calculate residuals, treating positive and negative residuals separately (Figure 7(D)). Working with residuals avoids the problem of biased averaging when positive and negative rates are treated separately or pooled as absolute values. The directional time series in

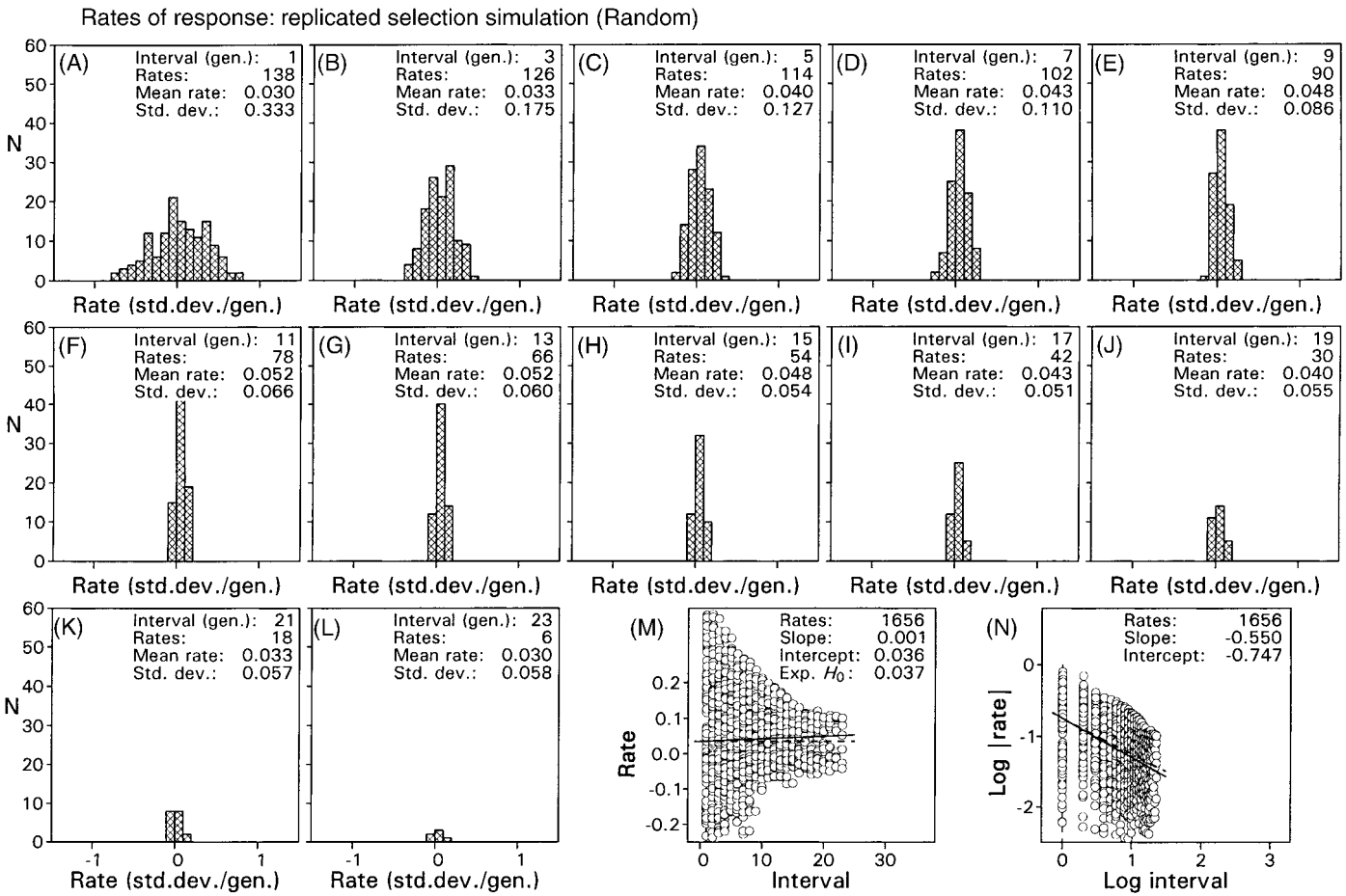

Figure 6. Comparison of rates of response in a 23-generation replicated generational-scale selection simulation involving random selection of rates each generation from a distribution with the mean and standard deviation for Falconer's control experiment (Figure 5(A)). (A-L) Histograms of rates calculated over different intervals of time (panels for even-numbered intervals are omitted to conserve space) – note the symmetry of mean rates in panels A and L, B and K, etc., while standard deviations decline systematically. (M) Rate versus interval [RI] plot combining rates for all time scales – predicted H_0 is value of regression for an interval of one generation. (N) Log absolute value of rate versus log interval [LRI] plot. Note that rates decline with interval length (solid line slope $= -0.550$) and almost match expectation for a random process (dashed line slope $= -0.500$).

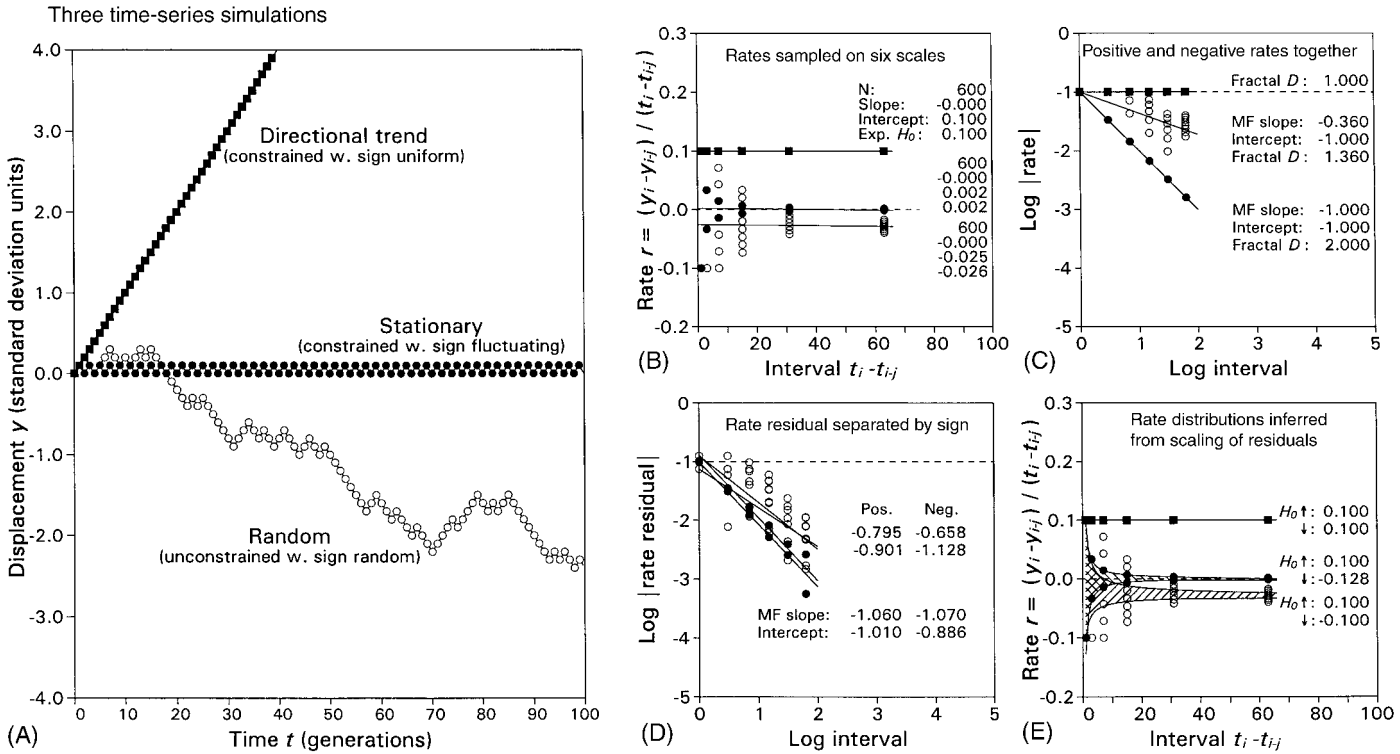

Figure 7. Three simulations showing interpretation of rate versus interval [RI] and log rate versus log interval [LRI] graphs in terms of model time series. (A) Solid squares in panel A show a directional trend with signs constrained to be uniformly positive, modeling persistent directional selection in successive generations. Solid circles in panel A show a stationary time series with signs constrained to fluctuate, modeling frequent reversals of direction tracking a stable adaptive peak. Open circles in panel A show a random time series with the sign at each step independent of that in previous steps. All three time series are gradual, changing at 0.1 standard deviation units per generation. (B) Distributions of rates calculated for each time series are based on six samples of 100 rates each, drawn at random from each time series on scales of 2, 4, 8, 16, 32, and 64 generations. (C) Ordinary LRI graph of rate distributions in panel B showing how each model time series can be distinguished by its fractional or fractal dimension D derived from the slope of the LRI plot. (D) Alternative LRI plot of rate residuals separated by sign. (E) Modal distributions of rates inferred from the scaling of rate residuals in panel D. Fifty percent or more of observed rates are expected to lie outside each of the shaded envelopes. MF slopes and intercepts are derived using a robust median fitting routine minimizing absolute deviations. Note that the directional time series is most constrained and hence has least variance. The random time series is least constrained and has greatest variance. All of the interpretative graphs (B–E) accurately recover the 0.1 standard deviation per generation rates of the model time series on a generational scale of time, although these values exceed time-averaged rates on longer time scales for any but purely directional trends.

Figure 7(B) has no residuals because R is independent of I, so it is not considered further here. The stationary time series in Figure 7(B) yields the residuals plotted as solid circles in Figure 7(D). The slope of the positive residuals is -1.060 and the slope of negative residuals is -1.070, with intercepts being -1.010 and -0.886 respectively. These yield the narrow hatched envelope fit to the solid circles in Figure 7(E). The random time series in Figure 7(B) yields the residuals plotted as open circles in Figure 7(D). The slope of the positive residuals is -0.795 and the slope of negative residuals is -0.658, with intercepts being -0.901 and -1.128 respectively. These yield the broader hatched envelope fit to the open circles in Figure 7(E).

The hatched envelopes in Figure 7(E) are not confidence intervals in the ordinary sense, but they do tell us about the distribution of rates based on a sample in each case. Since positive and negative residuals are fit separately, each line describes the median of a distribution of residuals: half are expected to be above the line and half below it. Hence we can say that at least half, and possibly more, of all rates are expected to fall above or below each hatched envelope. More than a quarter of all rates will fall above the upper line of the envelope if some from the upper half of the distribution of negative residuals fall in the upper half of the distribution of positive residuals, and more than a quarter of all rates will fall below the lower line of the envelope if some from the lower half of the distribution of positive residuals fall in the lower half of the distribution of negative residuals. A confidence interval is usually an interval that is conservative in describing what is *included*, but in this case the interval is conservative in describing what is *excluded*.

In most studies two numbers are required to represent H_0. One is the median value of negative rate residuals on a generational time scale ($H_0 \downarrow$). This number is generally negative because the rate corresponding to the median of negative residuals is generally negative. The other is the median value of positive rate residuals on a generational time scale ($H_0 \uparrow$). This number is generally positive because the rate corresponding to the median of positive residuals is generally positive. If we consider the signs of $H_0 \downarrow$ and $H_0 \uparrow$ to be important, then the expected value of H_0 is the average of $H_0 \downarrow$ and $H_0 \uparrow$, which will be a number close to zero in most studies. However, if the signs are unimportant, then the expected value of H_0 is the average of $|H_0 \downarrow|$ and $|H_0 \uparrow|$, which will be a number close to both $|H_0 \downarrow|$ and $|H_0 \uparrow|$ in most studies.

Evolutionary rates: empirical evidence

Evolutionary rates are known from sources representing change in different settings and change on different scales of time. A sampling of these is shown in Figure 8, taken from a survey of evolutionary rates that is not complete, but nevertheless likely to be representative of rates in general. My purpose in making the comparison is to show that all studies on all scales of time yield consistent results.

One source of evolutionary rates is experimental. All 413 of the nonzero autonomous rates calculated over single generation intervals of time in the Falconer replicated laboratory selection experiments (Figures 3(A), 4(A), and 5(A)) are shown in the LRI graph of Figure 8(A). In addition, 12 rates from field selection experiments of Reznick et al. (1997), calculated over intervals of about 7, 13, and 19 generations, are shown as well. Regression of logs of the absolute values of all rates on logs of their associated intervals yields an LRI slope of -0.639, which is between the value expected for random change (-0.500) and that expected for a stationary time series (-1.000) – and closer to the former. All of the Reznick et al., rates are positive. Regression of the logs of all positive experimental rates on the logs of their associated intervals yields an LRI slope of -0.600 and an intercept of -0.619. If we assume that residuals of all rates over the long term will have the same sign as the rates themselves, then we can calculate an expected upper median of positive rates, which here is 0.240 haldanes ($10^{-0.619}$). There are no negative experimental autonomous rates calculated on time scales longer than one generation in this sample, so it is not yet possible to calculate an independent lower median of negative experimental rates.

Hendry and Kinnison (1999) summarized a large set of 'microevolutionary' rates, based on field and laboratory studies, including those of Reznick et al., included in Figure 8(A) as experimental rates. These too can be studied in an LRI context (Figure 8(B)). Taken together, regression of logs of the absolute values of all rates on logs of their associated intervals yields an LRI slope of -1.046, which is almost exactly that expected for a stationary time series (-1.000). Separate treatment of rates reported as positive and negative yields an upper median of positive rates of about 0.660 haldanes and a lower median of negative rates of about -0.486 haldanes. Some 50% of all rates are expected to lie outside this range, involving rates on the order of ± 0.5 or more stand-

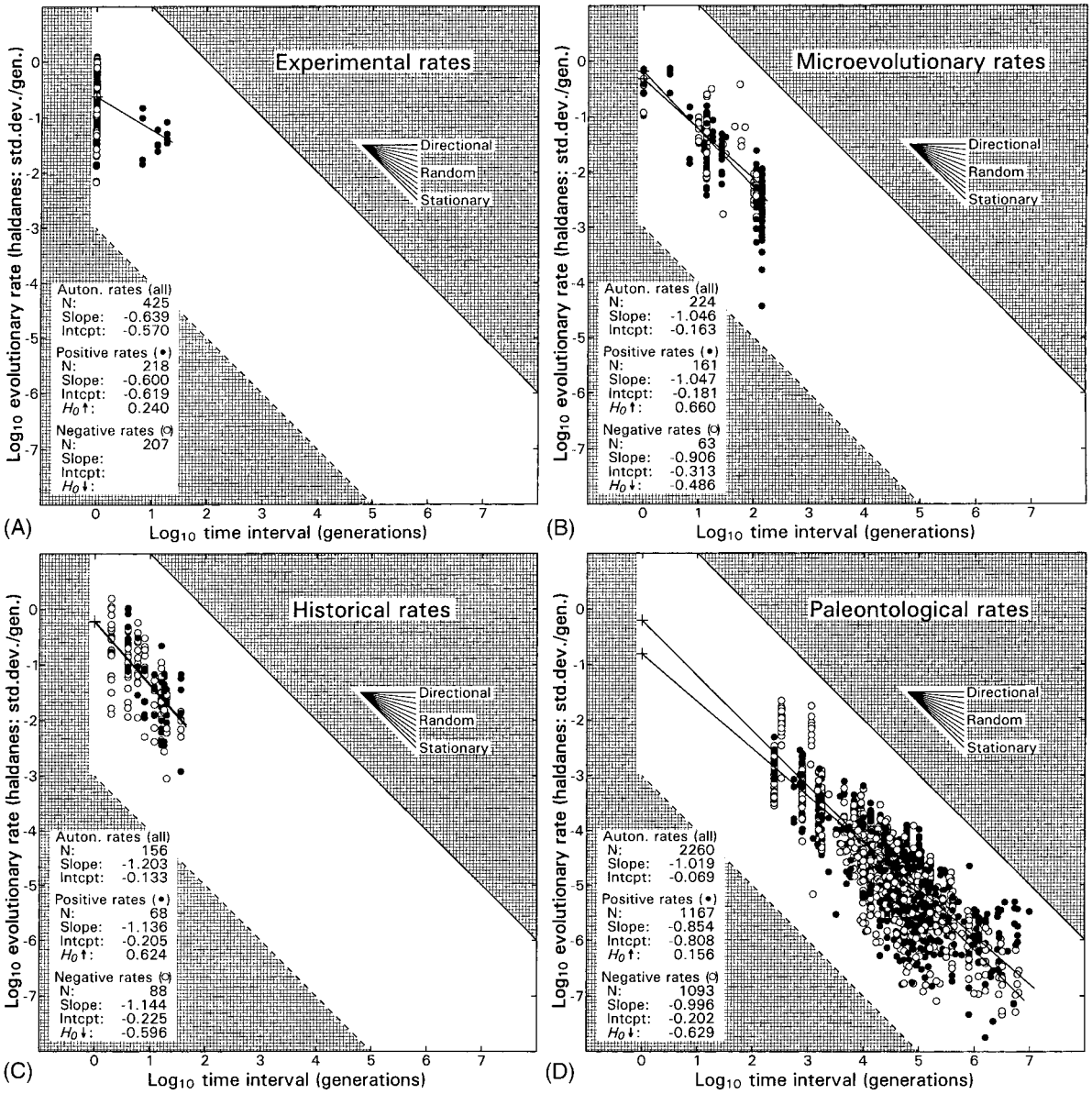

Figure 8. Temporal scaling of evolutionary rates from different sources calculated on different scales of time to show their comparability. (A) Laboratory selection experiments of Falconer (1973) and field selection experiments of Reznick et al. (1997). (B) Microevolutionary rates summarized in Hendry and Kinnison (1999). (C) Historical field study of Michigan mouse *Peromyscus gracilis* (unpublished study of L. Hester). (D) Paleontological studies of Colbert (1948), Haldane (1949), Maglio (1973), McDonald (1981), Malmgren et al. (1983), King and Saunders (1984), Flynn (1986), Lister (1989), Barnosky (1990), Forstén (1990), Geary (1990), Lich (1990), Gingerich (1991), Heaton (1993), Clyde and Gingerich (1994), Gingerich (1994), Gingerich and Gunnell (1995), Klein (1995), Gingerich (1996), Prothero and Heaton (1996), Polly (1997), Ruff et al. (1997), and Bloch and Gingerich (1998). All rates are autonomous rates, meaning that they are calculated over the shortest intervals spanning successive samples in a time series. Slopes and intercepts are derived using a robust median fitting routine minimizing absolute deviations. Note that some log rate intercepts are on the order of $10^{-0.2}$, yielding H_0 rates as high as about -0.6 or $+0.6$ standard deviations per generation. Hence we can expect rates on the order of 0.1–0.2 standard deviations to be common on the per-generation scale of the evolutionary process.

ard deviations per generation on a generational time scale.

A third study of rates of change in Michigan field mice collected during the past century is shown in Figure 8(C). Taken together, regression of logs of the absolute values of all rates on logs of their associated intervals yields an LRI slope of -1.203, which is even a little more negative than that expected for a stationary time series (-1.000). Treatment of positive and negative rates separately yields an upper median of positive rates of about 0.624 haldanes and a lower median of negative rates of about -0.596 haldanes. Some 50% of all rates are expected to lie outside this range, involving rates on the order of ± 0.6 or more standard deviations per generation on a generational time scale.

Finally, a large sample of paleontological rates collected from many sources is shown in Figure 8(D). Taken together, regression of logs of the absolute values of all rates on logs of their associated intervals yields an LRI slope of -1.019, which is again almost exactly that expected for a stationary time series (-1.000). Treatment of positive and negative rates separately yields an upper median of positive rates of about 0.156 haldanes and a lower median of negative rates of about -0.629 haldanes (the high intercept for negative rates is determined by short-term rates in two studies). Some 50% of all rates are expected to lie outside this range, involving rates on the order of ± 0.15 or more standard deviations per generation on a generational time scale.

We can now address the question posed in the introduction. Which studies – microevolutionary or macroevolutionary (or neither) – yield rates representative of the evolutionary process on a generational scale of time? Macroevolutionary studies yield rates on the order of 10^{-2}–10^{-6} haldanes calculated over intervals of geological time ranging from 10^2–10^6 generations. Microevolutionary studies yield rates on the order of 10^{-1}–10^{-2} haldanes calculated over intervals of historical time ranging from 10^1–10^2 generations. Microevolutionary and macroevolutionary studies are consistent in pointing to very high rates of change on a generational scale of time, although neither group is itself representative of change on this time scale. H_0 is 0.240 in panel A of Figure 8, -0.486 and $+0.660$ in panel B, -0.596 and $+0.624$ in panel C, and -0.629 and $+0.156$ in panel D. The most conservative H_0 values indicate change on the order of 0.2 standard deviations or more per generation on a generational time scale. These are evolutionary rates *on the time scale of the evolutionary process*. Rates from field studies and paleontological studies yield H_0 values that equal or exceed average rates from selection experiments like those shown in Figures 3 and 4. All evidence indicates that the evolutionary process is much more dynamic than we generally recognize.

Evolution appears slow when we consider net rates averaged over long intervals of time in the fossil record, but these are not rates on the time scale of the evolutionary process. Lande (1976, p. 333) found that change in fossil lineages can be explained by as few as about one selective death per million individuals per generation, and Lynch (1990, p. 739) found that rates of morphological change in fossil lineages are substantially below the minimum neutral expectation. Both results are surprising, but each study erroneously assumed that macroevolutionary rates calculated on geological scales of time represent evolution on the time scale of the process. All rates must be scaled against interval length for comparison on a common scale, and when this is done all studies yield consistent estimates of H_0.

Interpretation

If the process of evolution is so dynamic on a generational scale of time, why does it appear virtually stationary on longer scales of time? Here I will develop an example from Cenozoic mammal evolution as an illustration. This involves an heuristic evolutionary time-form lattice (Figure 9(A)) facilitating comparison of morphology to evolutionary time. Our observation that evolution takes place at rates on the order of 0.1 haldanes (rounding to the nearest order of magnitude) indicates that one 0.1 standard deviation unit of morphological difference is evolutionarily equivalent to one generation of temporal difference. Generations and 0.1 standard deviation units calibrate and scale the time-form lattice.

Now one of the smallest living mammals is the least shrew weighing about 3 or $e^{1.1}$ g, and the largest living mammal is the blue whale weighing about 100 metric tonnes or $e^{18.4}$ g. The standard deviation of body weight in mammals is about 0.15 units on a natural logarithmic scale. Hence the largest and smallest mammals living today differ by approximately 100 standard deviations, which is 10^3 0.1-standard-deviation units. These are physiological limits and mammals have never been much smaller or much larger: thus the time-form lattice for mammalian evolution is about 10^3 units wide.

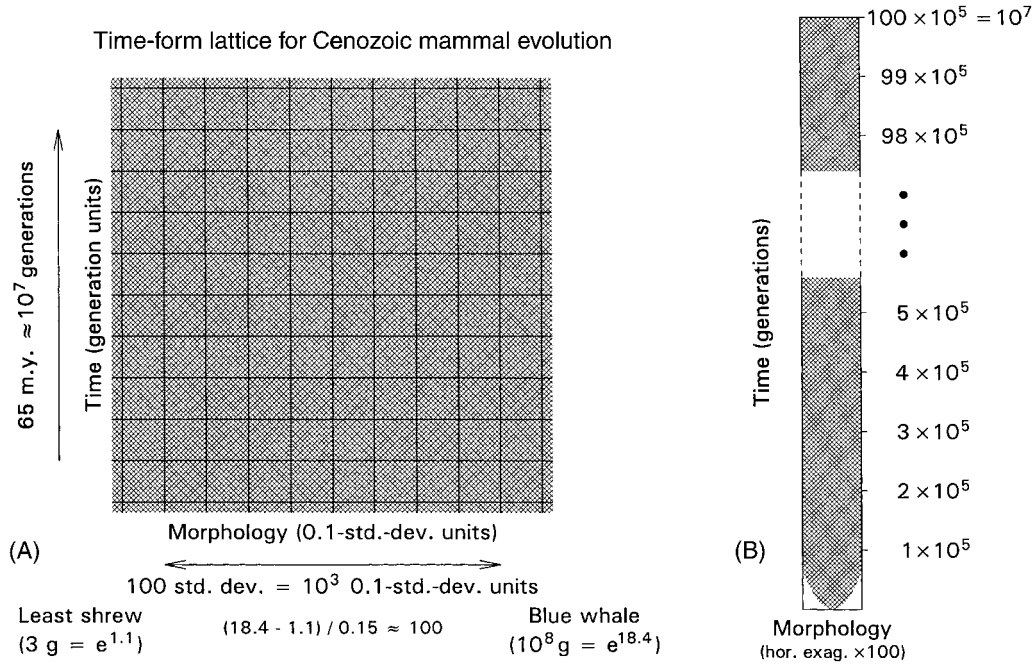

Figure 9. Heuristic time-form lattice to explain how evolution can be so dynamic in the short term, with H_0 rates on the order of 0.1 standard deviations per generation on the scale of one generation, and so static over longer intervals. (A) If H_0 rates are on the order of 0.1 standard deviations per generation, then evolution takes place on a time-form lattice where a 0.1 standard deviation step or difference in morphology corresponds roughly to a one-generation step or difference in evolutionary time. The smallest living mammal is the least shrew weighing about 3 or $e^{1.1}$ g, and the largest living mammal is the blue whale weighing about 100 metric tonnes or $e^{18.4}$ g. The standard deviation of body weight in mammals is about 0.15 units on a natural logarithmic scale. Hence the largest and smallest mammals living today differ by approximately 100 standard deviations or 10^3 0.1-standard-deviation units. These are physiological limits and mammals have never been much smaller or much larger: thus the time-form lattice for mammalian evolution is about 10^3 units wide. The generation time for an average living mammal is on the order of one year, and the Cenozoic history of the modern orders of mammals as we know them goes back 55–65 million years, which is conservatively about 10 million or 10^7 generations. Thus the time-form lattice for Cenozoic mammal evolution is about 10^7 units long temporally. (B) The lattice is not square, but some four orders of magnitude longer temporally than it is wide in form. Mammals starting at some average size at the beginning of the Cenozoic can be expected to have diffused and filled the lattice in less than 10^5 generations $[(500/1.96)^2 \approx 65000$ generations] – less than one percent of their subsequent Cenozoic history. Then they were constrained to evolve within the lattice for the remaining 99 percent of their history. Rates of evolution *on the time scale of the process* are so high that lineages rapidly find and fill most niches within their physiological limits. Then they change little until the system is perturbed.

The generation time for an average living mammal can be assumed to be on the order of one year, and the Cenozoic history of the modern orders of mammals, mammals as we know them, goes back 55–65 million years, which is conservatively about 10 million or 10^7 generations. Thus the time-form lattice for Cenozoic mammal evolution is about 10^7 units long temporally. The width and length of the lattice differ by four orders of magnitude!

The diagram in Figure 9(B) is designed to illustrate what a four-order-of-magnitude difference between morphology and time means for the history of mammals. If we assume that mammals start at an intermediate size and speciate rapidly enough to diffuse randomly across the time-form lattice at the rate of their morphological evolution, 0.1 standard deviations per generation, it will take some unknown number of generations for the diffusion to reach the edges of the lattice. At any given time t some 95% of all species will lie within N lattice units of the starting size, where N is given by:

$$N = 1.96 \cdot t^{0.5}. \tag{4}$$

Solving this for t, we can expect the whole lattice to be filled within about $(500/1.96)^2 \approx 65000$ generations. This is less than 10^5 generations, and less than 1% of the time available, as shown by expansion of the shaded envelope at the base of Figure 9(B). The diffusion could be even more rapid if species interact and selection plays a role. Empirical support for rapid filling of the Cenozoic mammal time-form lattice comes from Rose (1981) and Alroy (1998).

Once the lattice is filled, evolutionary time series can persist but become effectively stationary. This would be true of a single isolated lineage free to diffuse across an empty lattice through time (because the lattice is so much longer than it is wide), but it will be even more true when the lattice is filled with competing species. Calculation that the time-form lattice for body size in Cenozoic mammals is four orders of magnitude longer temporally than it is wide morphologically does not mean this is true for all morphological characteristics of mammals nor for all characteristics of other organisms. However, morphology appears generally to be more constrained for the simple reason that evolutionary history has been so very long in relation to the generation times of organisms.

Discovery that diffusion at documented rates can fill a whole evolutionary lattice in less than 1% of the time available for evolution is reminiscent of Gould's (1982, p. 84) characterization of geologically instantaneous punctuated change in branching speciation. He defined punctuated change as encompassing '1 percent or less of later existence in stasis'. The high rates of evolutionary change on a generational scale of time documented above, coupled with a temporally-long and morphologically-narrow time-form lattice for evolution provide evidence both permitting rapid 'punctuated' change and favoring long-term 'stasis' for all lineages that persist for any substantial number of generations. This result is consistent with Lynch's (1990) conclusion that stabilizing selection is a predominant evolutionary force keeping the long-term diversification of lineages well below its potential. Punctuated evolution is not an alternative to gradualism as proposed by Eldredge and Gould (1972), but rather results from the very dynamism of step-by-step gradual change on the generation-to-generation time scale of the evolutionary process. Punctuated patterns in the fossil record are real, but all that I am familiar with can be explained by natural selection on a generational time scale. Punctuated patterns do not seem to require any special mechanisms or processes beyond natural selection.

In this view, the long-term history of life is almost wholly decoupled from the evolutionary potential of high rates of change on a generational time scale. We would not know this if the history of life were short, but life's history, even the 65 m.y. history of Cenozoic mammals, is long in relation to the variation we can see and measure in the dimension(s) of form: form is clearly more constrained than time.

If the long-term history of life is predominantly a history of stasis, why are there so many shorter intervals of rapid change? One possibility, for which there is much evidence, is that the environmental and ecological 'theater' of the evolutionary play is never stable. Asteroids impact the earth, crustal plates move, our orbit around the sun changes on various time scales, volcanos erupt, sea level rises and falls, atmospheric gases accumulate and dissipate, oceanic circulation oscillates, climates change, floras are affected, faunas respond, and there are connections, interactions, and feedbacks at many levels. Thus evolution on the time-form lattice is started and restarted episodically. The history of life represented in subdivisions of the geological time scale reflect this. Phanerozoic time is divided into Paleozoic, Mesozoic, and Cenozoic eras, and the latter is subdivided into Paleocene, Eocene, etc., epochs, with each 'chapter' in the history of life being an interval occupied by important new macroevolutionary patterns of diversification.

We can speak of microevolutionary and macroevolutionary change, of microevolutionary and macroevolutionary patterns, and of history on microevolutionary and macroevolutionary scales, but it is not clear that there are any separate microevolutionary or macroevolutionary processes. Generational change is so rapid that stability comes soon, but the stability is often perturbed and then new microevolutionary and macroevolutionary patterns begin. Evolution on a generational scale, by itself, cannot explain microevolutionary change over multiple generations, nor can it explain macroevolutionary change over geological time. Each must be studied on its own time scale, remembering that the scale itself is important when relating one to another.

Acknowledgements

I thank D.S. Falconer for providing original data from his replicated selection experiments, and Laurel Hester for allowing me to use data from her unpublished field study of *Peromyscus gracilis*. Daniel C. Fisher, Michael Foote, H. David Sheets, and the editors provided substantial reviews improving the text.

References

Alroy, J., 1998. Cope's rule and the dynamics of body mass evolution in North American fossil mammals. Science 280: 731–734.

Barnosky, A.D., 1990. Evolution of dental traits since latest Pleistocene in meadow voles (*Microtus pennsylvanicus*) from Virginia. Paleobiology 16: 370–383.

Bloch, J.I. & P.D. Gingerich, 1998. *Carpolestes simpsoni*, new species (Mammalia, Proprimates) from the late Paleocene of the Clarks Fork Basin, Wyoming. Contributions from the Museum of Paleontology, University of Michigan 30: 131–162.

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

Colbert, E.H., 1948. Evolution of the horned dinosaurs. Evolution 2: 145–163.

Darwin, C., 1859. The Origin of Species. John Murray, London. Reprinted 1964 by Harvard University Press, Cambridge.

Dobzhansky, T., 1937. Genetics and the Origin of Species. Columbia University Press, New York.

Eldredge, N. & S.J. Gould, 1972. Punctuated equilibria: an alternative to phyletic gradualism, pp. 82–115 in Models in Paleobiology, edited by T.J.M. Schopf. Freeman, Cooper and Company, San Francisco.

Falconer, D.S., 1973. Replicated selection for body weight in mice. Genetical Research, Cambridge 22: 291–321.

Flynn, L.J., 1986. Species longevity, stasis, and stairsteps in rhizomyid rodents, pp. 273–285 in Vertebrates, Phylogeny, and Philosophy, edited by K.M. Flanagan & J.A. Lillegraven. University of Wyoming Contributions to Geology, Special Paper, Laramie.

Foote, M., 1994. Temporal variation in extinction risk and temporal scaling of extinction metrics. Paleobiology 20: 424–444.

Forstén, A.-M., 1990. Dental size trends in an equid sample from the Sandalja II cave of northwestern Yugoslavia. Paläontologische Zeitschrift, Stuttgart 64: 153–160.

Geary, D.H., 1990. Patterns of evolutionary tempo and mode in the radiation of *Melanopsis* (Gastropoda; Melanopsidae). Paleobiology 16: 492–511.

Gingerich, P.D., 1983. Rates of evolution: effects of time and temporal scaling. Science 222: 159–161.

Gingerich, P.D., 1991. Systematics and evolution of early Eocene Perissodactyla (Mammalia) in the Clarks Fork Basin, Wyoming. Contributions from the Museum of Paleontology, University of Michigan 28: 181–213.

Gingerich, P.D., 1993. Quantification and comparison of evolutionary rates. Am. J. Sci. 293A (Ostrom volume): 453–478.

Gingerich, P.D., 1994. New species of *Apheliscus*, *Haplomylus*, and *Hyopsodus* (Mammalia, Condylarthra) from the late Paleocene of southern Montana and early Eocene of northwestern Wyoming. Contributions from the Museum of Paleontology, University of Michigan 29: 119–134.

Gingerich, P.D., 1996. Rates of evolution in divergent species lineages as a test of character displacement in the fossil record: tooth size in Paleocene *Plesiadapis* (Mammalia, Proprimates), pp. 193–204 in Paléobiologie et Evolution des Mammifères Paléogènes: Volume Jubilaire en Hommage à Donald E. Russell, edited by M. Godinot & P.D. Gingerich. Palaeovertebrata, Montpellier.

Gingerich, P.D., 2000. Arithmetic or geometric normality of biological variation: an empirical test of theory. J. Theor. Biol. 204: 201–221.

Gingerich, P.D. & G.F. Gunnell, 1995. Rates of evolution in Paleocene-Eocene mammals of the Clarks Fork Basin, Wyoming, and a comparison with Neogene Siwalik lineages of Pakistan. Palaeogeog. Palaeocl. Palaeoecol. 115: 226–247.

Gould, S.J., 1982. The meaning of punctuated equilibrium and its role in validating a hierarchical approach to macroevolution, pp. 83–104 in Perspectives on Evolution, edited by R. Milkman. Sinauer Associates, Sunderland, Massachusetts.

Haldane, J.B.S., 1949. Suggestions as to quantitative measurement of rates of evolution. Evolution 3: 51–56.

Heaton, T.H., 1993. The Oligocene rodent *Ischyromys* of the Great Plains: replacement mistaken for anagenesis. J. Paleontol. 67: 297–308.

Hendry, A.P. & M.T. Kinnison, 1999. The pace of modern life: measuring rates of contemporary microevolution. Evolution 53: 1637–1653.

King, J.E. & J.J. Saunders, 1984. Environmental insularity and the extinction of the American mastodont, pp. 315–339 in Quaternary Extinctions: A Prehistoric Revolution, edited by P.S. Martin & R.G. Klein. University of Arizona Press, Tucson.

Klein, R.G., 1995. The Tor Hamar fauna, pp. 405–416 in Prehistoric Cultural Ecology and Evolution: Insights from Southern Jordan, edited by D.O. Henry. Plenum, New York.

Lande, R., 1977. On comparing coefficients of variation. Systematic Zoology 26: 214–217.

Lande, R., 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30: 314–334.

Lande, R. & S.J. Arnold, 1983. The measurement of selection on correlated characters. Evolution 37: 1210–1226.

Lerman, A., 1965. On rates of evolution of unit characters and character complexes. Evolution 19: 16–25.

Lewontin, R.C., 1966. On the measurement of relative variability. Systemat. Zool. 15: 141–142.

Lich, D.K., 1990. *Cosomys primus*: a case for stasis. Paleobiology 16: 384–395.

Lister, A.M., 1989. Rapid dwarfing of red deer on Jersey in the last interglacial. Nature 342: 539–542.

Lynch, M., 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. Am. Natural. 136: 727–741.

Maglio, V.J., 1973. Origin and evolution of the Elephantidae. Transactions of the American Philosophical Society 63: 1–149.

Malmgren, B.A., W.A. Berggren & G.P. Lohmann, 1983. Evidence for punctuated gradualism in the late Neogene *Globorotalia tumida* lineage of planktonic foraminifera. Paleobiology 9: 377–389.

Mandelbrot, B.B., 1967. How long is the coast of Britain? Statistical self-similarity and fractional dimension. Science 156: 636–638.

Mandelbrot, B.B., 1983. The Fractal Geometry of Nature. W.H. Freeman, San Francisco.

McDonald, J.N., 1981. North American Bison: Their Classification and Evolution. University of California Press, Berkeley.

McShea, D.W. & D.M. Raup, 1986. Completeness of the geological record. J. Geol. 94: 569–574.

Polly, P.D., 1997. Ancestry and species definition in paleontology: a stratocladistic analysis of Paleocene-Eocene Viverravidae (Mammalia, Carnivora) from Wyoming. Contributions from the Museum of Paleontology, University of Michigan 30: 1–53.

Prothero, D.R. & T.H. Heaton, 1996. Faunal stability during the early Oligocene climatic crash. Palaeogeo. Palaeocl. Palaeoecol. 127: 257–283.

Reznick, D.N., F.H. Shaw, F.H. Rodd & R.G. Shaw, 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). Science 275: 1934–1937.

Roff, D.A., 1997. Evolutionary Quantitative Genetics. Chapman and Hall, New York.

Rose, K.D., 1981. Composition and species diversity in Paleocene and Eocene mammal assemblages: an empirical study. J. Verteb. Paleontol. 1: 367–388.

Ruff, C.B., E. Trinkaus & T.W. Holliday, 1997. Body mass and encephalization in Pleistocene *Homo*. Nature 387: 173–176.

Sadler, P.M., 1981. Sediment accumulation rates and the completeness of stratigraphic sections. J. Geol. 89: 569–584.

Sadler, P.M., 1993. Time scale dependence of the rates of unsteady geologic processes, pp. 221–228 in Rates of Geologic Processes: Tectonics, Sedimentation, Eustasy, and Climate. Implications for Hydrocarbon Exploration, edited by J.M. Armentrout, R. Bloch, H.C. Olson & B.F. Perkins. Gulf Coast Section, SEPM Foundation: 14th Annual Research Conference, Baton Rouge, Louisiana.

Sokal, R.R. & F.J. Rohlf, 1981. Biometry. W.H. Freeman, San Francisco, 2nd edn.