

**Rates of evolution in Plio-Pleistocene
mammals: six case studies**

PHILIP D. GINGERICH

In :

**Morphological change in Quaternary mammals
of North America**

Edited by

ROBERT A. MARTIN
Murray State University

ANTHONY D. BARNOSKY
University of California at Berkeley

1993

6

Rates of evolution in Plio-Pleistocene mammals: six case studies

PHILIP D. GINGERICH

Evolution, in general, means change over time: The end is different from the beginning. Change is not inevitable, but it seems to be common. And some change can almost always be observed, given enough time. When do we see change? How much time does it take? To study evolution we have to measure change and time, and to understand evolution we have to know how the two are related.

Evolution in biology is often described as a process, but it is really a collection of processes – including the processes of mutation, dispersion, drift, and selection – that together produce whatever change we see. These processes may each be complicated individually, and their interactions produce additional complexity. Evolution is widely acknowledged as a fundamental concept in biology, and yet we have a surprisingly anecdotal and casual knowledge of how evolutionary change takes place. One way to improve this understanding is to quantify change over time in terms of rate. Evolutionary rates are quantitative expressions relating change to time.

Evolution takes place on many scales of time. Field and laboratory experiments can be designed to study change on short time scales, and fossils provide the most direct and best information about evolution on long time scales. The principal problem with the fossil record is that the time scales involved, typically millions of years, are so long that they are difficult to relate to the time scales of our lifetimes (and those of other organisms). Biologists as a group have a surprisingly poor understanding of evolution on a geological scale of time, and paleontologists as a group have a surprisingly poor understanding of evolution on a biological scale of time. One reason for this is that we have almost no

record of changes on intermediate scales of time, scales of hundreds or thousands of years, that would permit evolution on a laboratory scale of time to be related to evolution on a geological scale.

Fortunately biologists can sometimes study evolution on a scale of hundreds or thousands of years by taking advantage of accidents and events that have affected organisms, (e.g., chance colonization events) and have been documented in human history. And in favorable circumstances, paleontologists, too, can resolve time on scales of hundreds or thousands of years. It is generally true that time is easier to resolve the less far back we are in geological time, and the Pleistocene, spanning much of the past 2 million years, is important in this regard. The six Plio-Pleistocene studies discussed here illustrate a paleontological approach to understanding evolution on a wide range of time scales.

Quantification of evolutionary rates

Quantification of evolutionary rates is still in its infancy, and the history of ideas on this can be summarized rather easily. Time is uniquely important in paleontology, and it comes as no surprise that paleontologists figure prominently in the quantification of evolutionary rates. George Gaylord Simpson provided the first substantial treatment of rates in his book *Tempo and Mode in Evolution* (1944). At that time the geological time scale was poorly calibrated numerically, and Simpson was unwilling to trust estimates of the durations of epochs. He published a table of various measures of tooth size in horses and showed graphically that if the length of the ectoloph evolved at a constant rate, then the height of the paracone did not (Simpson, 1944, Figure 2). Simpson also provided a graph of these two characteristics plotted against geological time, showing that in all likelihood neither had evolved at a constant rate (Simpson, 1944, Figure 4). He noted that the slopes of lines in his graphs were proportional to rates of evolution, but did not attempt to quantify these. Simpson's most noteworthy contribution to quantification was recognition from the beginning that proportional change, not absolute change, is the quantity of interest, and consequently that a proportional (logarithmic) scale is the appropriate measurement scale for morphology.

J. B. S. Haldane (1949) quantified evolutionary rates in terms of proportional change and proposed the *darwin*, defined as change by a factor of e per million years (where e is the base of natural logarithms), as a convenient rate unit. Haldane calculated rates of evolution in horses

for Simpson's two linear measures and for their ratio (Haldane's Table 1), but seems not to have recognized the incomparability of rates quantified for measures of different dimensions. Haldane mentioned the desirability of measuring morphological change in standard deviation units, and he also mentioned the desirability of measuring time in generations, suggestions developed further in this chapter.

Björn Kurten (1959) measured rates of evolution in various ways in mammals that lived during the pre-Pleistocene (Tertiary), the Pleistocene, and the post-Pleistocene (Holocene), finding that each time period had its own characteristic range. Pre-Pleistocene mammals had rates ranging from 0.003 to 0.2 darwins (mean 0.02 *d*), Pleistocene mammals had rates ranging from 0.12 to 2.3 *d* (mean 0.5 *d*), and post-Pleistocene mammals had rates ranging from 3.7 to 43 *d* (mean 12.6 *d*). This led Kurten to suggest that rates for Quaternary (Pleistocene and post-Pleistocene) mammals may have been higher than those for Tertiary (pre-Pleistocene) mammals, because environmental change in the Quaternary was rapid and revolutionary, whereas environmental change in the Tertiary was slow and gradual. In addition, Kurten suggested the possibility that slow Tertiary rates might be partially or even wholly spurious because they are based on samples millions of years apart, with intervening histories that may have contained any amount of fluctuation at higher rates. Rates measured over long intervals of time are generally lower than rates measured over shorter intervals, and this is due in part to time-averaging of "fluctuating" higher rates (Gingerich, 1983). This does not mean that any are spurious, but simply that the effect of interval length must be considered when rates are compared.

Log rate-log interval graphs

Evolutionary rates do not exist independently of the interval over which they are measured. This observation is easily accommodated by plotting rates in an interval context. The graphs employed here have interval length plotted on a proportional scale on the abscissa, and the absolute value of evolutionary rate plotted on a proportional scale on the ordinate (\log_{10} is used for convenience, because 10 is familiar as the base of our numbering system, but any base could be used). In the following discussion these "log rate-log interval" graphs are referred to as LR-LI graphs. Absolute values of rates are plotted on the assumption that rates of proportional increase are no different than rates of proportional decrease, and assignment of direction is arbitrary in any case.

Processes

Any real process producing systematic change (e.g., natural selection for small body size) does so at some characteristic intrinsic rate that is independent of interval length, meaning that the simple result of a single real process carried out long enough will plot as a horizontal distribution of average slope zero (0.0) on a rate-versus-interval graph (and on an LR-LI graph). The result is a distribution, because intrinsic rates and net rates (discussed later) have variance, and the horizontal line labeled "Process" (for process change) in the inset key in the following LR-LI graphs represents the zero slope of this possibility.

Stasis

Alternatively, a simple process operating at zero rate, or two or more processes that cancel each other out (as well as absence of a process), will produce no systematic change over time (stasis), meaning that differences observed by chance will tend to be constant while interval length increases. Rates calculated from these necessarily will decrease in inverse proportion to interval length, and in this alternative case, if carried out long enough, rates will plot as a distribution with an average slope of negative unity (-1.0) on an LR-LI graph. Again, the result is a distribution, because intrinsic and net rates have variance (lacking variance, stasis at zero rate could not be plotted on an LR-LI graph). The line of slope -1.0 labeled "Stasis" in the inset key of the following LR-LI graphs represents the slope of this possibility.

Pure process and pure stasis are two extreme possibilities in comparing rates measured over a range of time intervals, but neither necessarily happens all the time. Process and stasis can be combined in any proportion, and the slope of the resulting rate distribution on an LR-LI graph will reflect the proportion. The slope will be near zero if process predominates, and near -1.0 if stasis predominates.

Random change

One additional possibility, random change, deserves mention. Appearance of randomness usually is taken to mean that real component factors or component processes interact in complex ways that defy separation. For some purposes, these can be adequately modeled together rather than studied individually. Random walks can mimic

process, and they can mimic stasis, but carried out long enough most will plot as a distribution with a slope of -0.5 on an LR-LI graph. The line of slope -0.5 labeled "Random" in the inset key of the following LR-LI graphs represents the slope of this possibility. Random change is an intermediate null model against which both process change and stasis must be compared to evaluate their possible significance (Gingerich, 1992).

Intrinsic rates and net rates

All stepped processes (including random drift) have two rates: an *intrinsic rate* that is the average rate observed at each step of the process (the generation-by-generation rate in evolution), and a *net rate* that is the average rate calculated over one or more steps (generations). The net rate can never exceed the intrinsic rate, although net rates approximate intrinsic rates in any real process. Net rates typically decrease in proportion to the square root of step number for random processes (hence the -0.5 slope on an LR-LI graph), and net rates decrease in proportion to step number in stasis (hence the -1.0 slope on an LR-LI graph).

Empirical limits

Empirically it turns out that rates of evolution have an upper bound and a lower bound, and both decrease systematically with interval length. Log rate of the upper bound and log rate of the lower bound decrease in inverse proportion to log interval, and each plots as a straight line of -1.0 slope. On an LR-LI graph, the upper bound is shown as a solid line because it reflects both a structural limit and an artifact of limited perception, and the lower bound is shown as a dotted line because it is purely an artifact of limited measurement.

The empirical upper and lower bounds of evolutionary rates define the limits of a broad distribution of possible rates having the -1.0 slope characteristic of stasis, which indicates that stasis, rather than process change or drift, predominates in long-term evolution.

Temporal scaling of evolutionary rates

The inverse relationship of evolutionary rates to the interval of time over which they are calculated means that rates can be compared directly

only if they are calculated over intervals of the same length. When a distribution of rates calculated over a range of intervals is available, it is sometimes possible to use the relationship of rate to interval within this distribution to predict rates for some interval outside the original range of intervals. This *temporal scaling* is useful when comparing different distributions, and it is particularly useful for predicting intrinsic rates on a scale of one generation or a scale of 1 year.

Confidence on other time scales

Inference concerning the relationship of evolutionary rates and time intervals is most easily carried out in an LR-LI context. Log rate is a derived variable dependent on log interval; however, ordinary least-squares regression cannot be used for extrapolation and prediction in this context because distributions of derived log rates are usually negatively skewed, which means that medians are better than mean values as measures of location and that squaring deviant outliers would bias regression. The method of slope-and-intercept computation employed here is *robust maximum likelihood* (RML) estimation minimizing absolute deviations. The distribution of residual variation about a line with the computed slope and intercept is shown as an inset histogram in the lower left corner of all LR-LI graphs. Confidence intervals are calculated by bootstrapping, following Efron and Tibshirani (1986). A linear model is used for extrapolation and prediction because of its simplicity, recognizing that any relationship of log rate to log interval may not be linear over its entire range.

Rate units

Haldane (1949) proposed a standard unit for rates of morphological evolution called the *darwin*, quantifying rates of morphological evolution in terms of factors of e and time in millions of years:

$$\text{rate } (d) = \frac{\ln x_2 - \ln x_1}{t_2 - t_1}, \quad (6.1)$$

where $\ln x_1$ and $\ln x_2$ are individual natural log (ln) measurements or sample means of ln measurements at times t_1 and t_2 , respectively. This is useful when comparing changes in measures of the same dimension (linear, areal, or volumetric) per unit time, but it is not generally useful when comparing changes in length with changes in area, or changes in

length with changes in volume, or changes in length with changes in any compound ratio (shape).

Haldane mentioned the desirability of measuring morphological change in standard deviation units, and our colleague Björn Malmgren has suggested a rate unit called the *simpson*, quantifying rates of morphological evolution in terms of standard deviations per year or per million years. This has the advantage that standard deviations are measured in the original measurement units, making rates in standard deviation units independent of original dimension and thus more widely comparable.

Haldane also mentioned the desirability of measuring time in generations, and another rate unit of general interest is what I call the *haldane* (*h*), quantifying rates of morphological evolution in terms of standard deviations per generation. Morphological change still has to be corrected for proportion, and we are thus interested in change in natural logarithms (ln) of measured quantities, expressed in standard deviations of the natural log measurements. That is,

$$\text{rate } (h) = \frac{(\ln x_2 - \ln x_1) / s_{\ln x}}{t_2 - t_1} = \frac{z_2 - z_1}{t_2 - t_1}, \tag{6.2}$$

where *x* and *t* are defined as in equation (6.1), and $s_{\ln x}$ is the pooled standard deviation of $\ln x_1$ and $\ln x_2$. Natural logarithms are convenient here because the coefficients of variation commonly reported in the literature are a good approximation of the standard deviation of ln measurements (Lewontin, 1966). The term $\ln(\text{mean } x)$ is used to approximate the mean of $\ln x$ when necessary. The notation h_0 refers to intrinsic evolutionary rate in haldanes estimated at a scale of one generation, that is, standard deviations per generation at a scale of one generation (which is simply the rate intercept where log interval = 0 on any LR-LI graph).

Generation time (*G*) is related to body mass (*M*) in mammals by the power function

$$G = 0.16M^{0.36}, \tag{6.3}$$

where *G* is measured in years, and *M* is measured in grams. This equation is derived by regression of log *G* on log *M* ($N = 66$, $r = 0.82$), and it can be used to estimate generation time from body mass when body mass is known. Data used in the regression come from Eisenberg (1981), with generation time taken as the sum of age at first mating plus gestation time. Equation (6.3) yields an approximation that should be rounded

downward for “*r*-selected” mammals and upward for “*K*-selected” mammals, especially those living in seasonal climates.

Rates in Plio-Pleistocene mammals: six case studies

Strait Canyon Microtus

The microtine rodent *Microtus pennsylvanicus* (meadow vole) was a common constituent of Pleistocene mammalian faunas of North America (and it is still widely distributed today). A. Barnosky studied samples of *M. pennsylvanicus* from 7 of 46 sampled intervals in Strait Canyon Fissure in Highland County, Virginia (Barnosky, 1990, and personal communication). Two of these sampling intervals 1.35 m apart have radiocarbon ages of 18,420 and 29,870 ky BP. Interpolation and extrapolation suggest that the seven *Microtus*-bearing intervals range in age from about 47,680 ky BP to about 20,960 ky BP. Samples are as few as 1,270 ky or as many as 26,720 ky apart in time. The *M. pennsylvanicus* weigh about 30 g and have a life span of about 1 year. Substitution in equation (6.3) indicates that *Microtus* probably reproduced at a rate of about three generations per year in the Pleistocene.

Rates of change of the length of the last upper molar (M^3) in Strait Canyon *Microtus* are shown in the LR-LI graph in Figure 6.1. Twenty-one rates can be calculated for six different intervals of time ranging from 3,810 generations to 80,160 generations. Twenty of the rates are nonzero. Thirteen rates are positive, indicating evolution toward larger size, and seven are negative, indicating evolution toward smaller size. The median rate is 0.000016 standard deviation per generation ($10^{-4.798}$), on a median time scale of about 33,000 generations ($10^{4.519}$), numbers that mean little by themselves. Fortunately, the *M. pennsylvanicus* data contain additional information.

All of the *Microtus* rates taken together have an RML slope of $- .512$, which suggests random change over time. The mean absolute deviation is 0.245. There are few enough data points that a bootstrapped 95% confidence interval on the slope ranges from -0.116 to -1.250 , ruling out process change (slope 0.0), but including both randomness (slope -0.5) and stasis (slope -1.0). The RML intercept estimate is $h_0 = 0.003$ ($10^{-2.487}$), which suggests an intrinsic evolutionary rate of 0.003 standard deviation per generation on a time scale of one generation. Artificial selection experiments achieve intrinsic rates on the order of

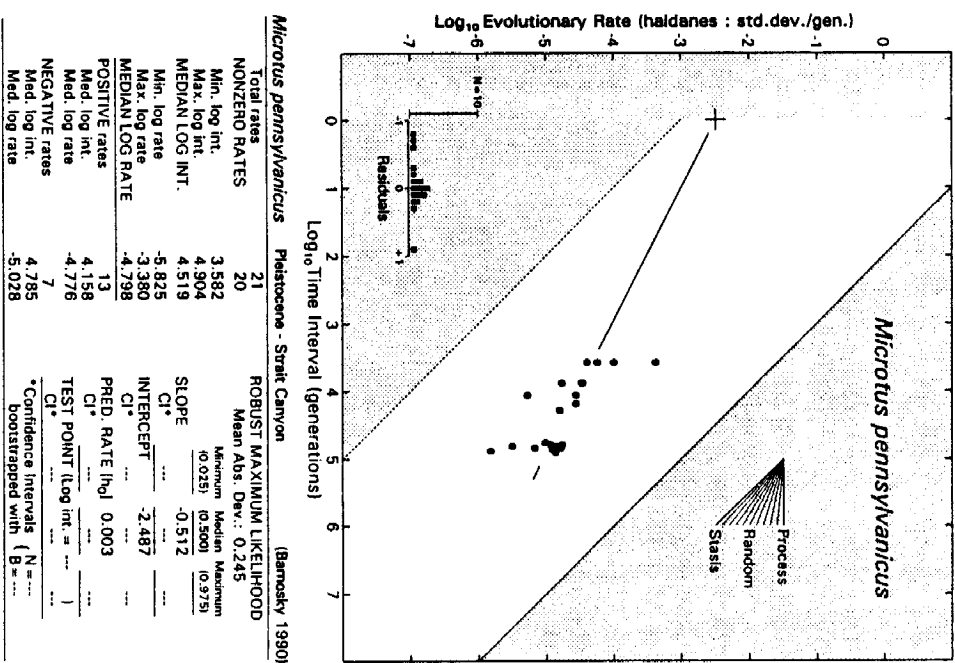


Figure 6.1. LR-LI graph for measurements of M^1 length in seven samples of late Pleistocene *Microtus pennsylvanicus* from Strait Canyon Frisure in Virginia. Solid circles are individual rates. Samples range in age from about 47,680 to 20,960 ky BP. *Microtus* is assumed to have had three generations per year, and rates are calculated over intervals ranging from 3,810 to 80,160 generations. Slope of rate distribution is -0.512 . Intercept is -2.487 , yielding an intrinsic rate $h_0 = 0.003$ standard deviation per generation on a time scale of one generation. This sample of rates is small enough that confidence intervals on the slope, extrapolated intercept, and h_0 (not shown) are very broad. Vertical scale bar represents $N = 10$ on inset histogram of residuals (mean absolute deviation is 0.245).

0.200 standard deviation per generation, and this *Microtus* result is about two orders of magnitude less than rates imposed by strong selection. However, here again there are few enough data points in the original rate sample that a bootstrapped 95% confidence interval for the predicted value of h_0 ranges from 0.000005 to 8.300 standard deviations per generation.

This example illustrates how net rates of evolution calculated on different time scales can be used to predict an intrinsic rate of change on a time scale of one generation, but the result cannot be distinguished from random change over time, and the confidence interval on the prediction is very broad.

Sandalia II Equus

A. Forsten (1990) studied late Pleistocene samples of the horse *Equus germanicus* from 7 of 12 sample intervals in the *Sandalia II* cave near Pula, in Istria, northwestern Yugoslavia. Two levels in the cave have radiocarbon ages, one near the bottom of the *Equus*-bearing sequence (25,340 ky BP), and one near the top (10,830 ky BP). Extrapolation and interpolation suggest that the seven *E. germanicus*-rich levels have ages of approximately 27,140, 24,740, 19,385, 14,675, 12,320, 11,575, and 10,830 ky BP. Successive samples differ by as few as 745 years and by as many as 5,355 ky, and *Equus* is assumed to have had a generation time of about 3 years in the Pleistocene.

Rates of change of measures of tooth size taken at the base of the crown and at the occlusal surface in *Sandalia II Equus* are shown in the LR-LI graph in Figure 6.2. A total of 714 rates can be calculated for intervals of time ranging from about 248 to 5,432 generations, and 666 of these rates are nonzero. Of the 666, 274 rates are positive, and 392 are negative, indicating a predominance of evolution toward smaller size. The median rate is 0.00021 standard deviation per generation ($10^{-3.686}$), on a median time scale of 2,466 generations ($10^{3.392}$). This median rate is about 13 times that for Strait Canyon *Microtus*, but the *Equus* rates are measured on a shorter median time scale.

All of the *Equus* rates taken together have an RML slope of -0.797 , which lies on the stasis side of random. Here there are enough data points to constrain confidence intervals, and bootstrapped confidence intervals for the slope, intercept, and predicted intrinsic rate h_0 are shown in Figure 6.3. The 95% confidence interval on the slope ranges from -0.731 to -0.894 . This narrow confidence interval rules out both

