

Phenotypic Evolution: The Ongoing Synthesis

The Faculty of Oregon State University has made this article openly available.
Please share how this access benefits you. Your story matters.

Citation	Arnold, S. J. (2014). Phenotypic Evolution: The Ongoing Synthesis. <i>The American Naturalist</i> , 183(6), 729-746. doi:10.1086/675304
DOI	10.1086/675304
Publisher	University of Chicago Press
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsfuse

Phenotypic Evolution: The Ongoing Synthesis

(American Society of Naturalists Address)*

Stevan J. Arnold[†]

Department of Integrative Biology, Oregon State University, Corvallis, Oregon 97331

Online enhancement: appendix.

ABSTRACT: I explore the proposition that evolutionary biology is currently in the midst of its greatest period of synthesis. This period, which I call the Ongoing Synthesis, began in 1963 and continues at the present time. I use analysis of citations, conduct, and content to compare the Ongoing Synthesis to widely recognized periods of synthesis in the nineteenth and twentieth centuries. To compare content, I focus on phenotypic evolution and compare current efforts with George Gaylord Simpson's struggle to understand evolution in deep geological time. The essence of current effort is captured by the question, What is the best model for phenotypic evolution? Although many investigators are actively engaged in answering this question, I single out two examples of my own collaborative work for emphasis here. These two studies share three important characteristics: diagnosis of evolutionary pattern using massive data sets, validation of model parameter values using compilations of estimates (e.g., heritability, stabilizing selection, distance to an intermediate optimum), and identification of evolutionary process using alternative models of stochastic evolution. Our primary findings (discovery of the blunderbuss pattern and the result that rare bursts of evolution carry lineages out of established adaptive zones) compare favorably with important insights from the Modern Synthesis.

Keywords: adaptive radiation, adaptive zone, quantum evolution, blunderbuss pattern, deep evolutionary time.

Introduction

When I set to work on this article, my mind was drawn back to a conversation I had more than 50 years ago. I was an undergraduate at the University of California at Berkeley, working as a curatorial assistant at the Museum of Vertebrate Zoology, when Ernst Mayr's book on animal speciation burst on the scene in 1963. Taking the ongoing buzz about the book for granted, I asked Professor Seth B. Benson, a senior curator of mammals, about the impact of Mayr's earlier book in 1942. Benson said that no one in the museum was much surprised by Mayr's conclusions;

* Presidential address to the American Society of Naturalists, July 2012, Ottawa, Canada.

[†] E-mail: arnolds@science.oregonstate.edu.

Am. Nat. 2014. Vol. 183, pp. 729–746. © 2014 by The University of Chicago. 0003-0147/2014/18306-54679\$15.00. All rights reserved.

DOI: 10.1086/675304

they mirrored contemporary understanding. Benson's insouciance may not have been a widespread reaction. Indeed, Mayr's 1942 and 1963 books are now widely regarded as iconic examples of synthesis in evolutionary biology. But looking beyond this particular incident, I wonder whether periods of extraordinary synthesis are generally experienced as ordinary science.

Are we now in the midst of one of the great synthetic periods in evolutionary biology? To get an answer for this essay, I compare two widely recognized periods of synthesis (1930–1932 and 1937–1950) with the contemporary scene in evolutionary biology. Citation statistics provide one basis for comparison. Another basis is provided by comparing George Gaylord Simpson's (1944) synthesis with current efforts to answer the question, What is the best model for phenotypic evolution? To develop this comparison, I focus primarily on two examples of synthesis from my own recent collaborative work. Both of these comparisons support the contention that we are indeed in the midst of a great synthetic period.

Synthesis in Evolutionary Biology: Then and Now

A modern synthetic period comparable to the great periods of the past is sometimes portrayed as desirable or anticipated but still out of reach rather than ongoing (Gould 1980; Pigliucci 2009). That anticipated synthesis is heralded because it would supplant outdated theory or at least extend existing theory (Pigliucci and Müller 2010; Futuyma 2011). One might also suppose that periods of great synthesis are times of harmony, during which all eyes are focused on the same prize. My analysis of the record does not support either position. Instead, I think the picture is one of continual synthesis, sometimes interrupted by periods of inertia but consistently rife with dissension and discord, and with synthesis accelerating at the present time. This view is especially supported by an examination of citation records over the past 150 years.

To quantify the impact of past and present synthetic volumes in evolutionary biology, I compiled citation counts for 59 influential synthetic books and articles (table

A1, available online). I judged works to be synthetic if they were integrative, book-length treatments that provided broad coverage of an important topic in evolutionary biology and influential if they had accumulated at least 500 citations in Google Scholar as of June 4–5, 2012 (see table A1 for additional details). The 59 works in table A1 represent a systematic sample rather than an exhaustive inventory. Furthermore, several limitations should be kept in mind when interpreting the graph derived from this sample: (1) Because Google Scholar is a product of the corporate world, the methods it uses to compile citations cannot be scrutinized. (2) No corrections were made for the number of years that have elapsed since publication or for the ever-expanding population of scientists who cite published articles. (3) Some important, meritorious works are not included in the sample. For example, Futuyma (1979) is inexplicably absent from Google Scholar's compilations. In other cases, I undoubtedly overlooked important works. Despite these limitations, a graphical analysis of citation provides a revealing supplement to the usual historical methods of diagnosing periods of synthesis.

Periods of evolutionary synthesis are recognizable as surges in citations followed by lulls. In figure 1, the cu-

mulative number of citations of 59 influential synthetic books is plotted against the year of publication. The longest-lasting lull (1890–1920) follows the publications of Darwin and Wallace during the period 1859–1881. Long-recognized synthetic periods occurred in 1930–1932 and 1937–1950. Both of these periods were followed by quiescent periods, 1932–1937 and 1950–1963. From the standpoint of a surge in publications, the period 1963–2004 could easily be classified as a synthetic period. When we consider the source of the contemporary surge, we find that it comes from the full range of topics and areas in evolutionary biology and not just from a few areas widely recognized as hot, such as genomics and evo-devo (see appendix, available online). The diagnosis of an extraordinary period is also strengthened by the realization that the publications in the period 1963–2004 will continue to accumulate citations and will probably do so at a faster rate than publications in the earlier periods. In other words, if we looked at this same graph (fig. 1) 50 years from now, we would expect the far-right portion to have increased considerably in elevation. Does this most recent 40-year interval also qualify as a period of synthesis on the basis of what was accomplished? To answer that ques-

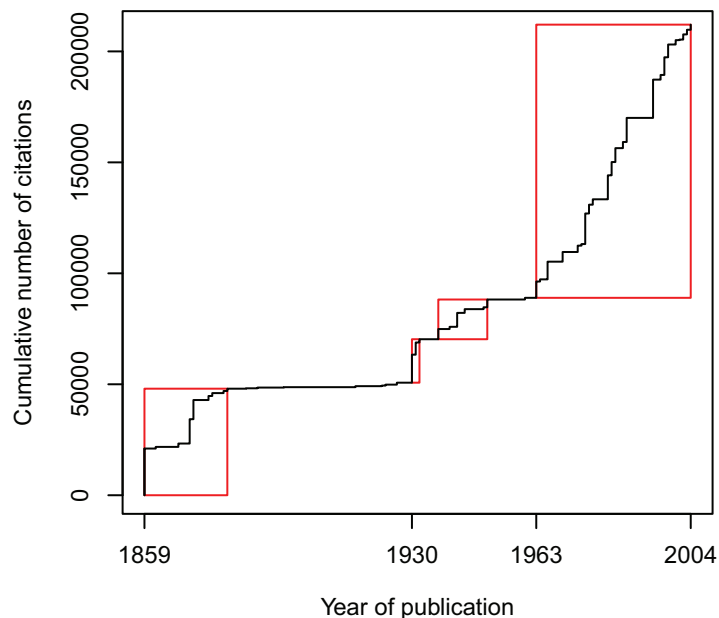


Figure 1: Cumulative number of citations of 59 influential synthetic works in evolutionary biology plotted against year of publication. Citation data were gathered using Google Scholar in June 2012. A list of the works and their citations is available in the online appendix. The four red rectangles show the citation contribution (height of rectangle) and temporal span (width of rectangle) of four synthetic periods: the Classic Period (Darwin and Wallace, 1859–1881), the Population Genetic Synthesis (Fisher, Wright, and Haldane, 1930–1932), the Modern Synthesis (Dobzhansky, Goldschmidt, Mayr, Huxley, Simpson, Schmalhausen, and Stebbins, 1937–1950), and the Ongoing Synthesis (31 authors publishing during the period 1963–2004). The Ongoing Synthesis accounts for more than half (55%) of the accumulated citations (212,068) in the sample of 59 influential works.

tion, I will first consider the content of the two earlier synthetic periods.

In the period 1930–1932, Ronald A. Fisher's, Sewall Wright's, and J. B. S. Haldane's publications are a bouquet that captured the blossoming of population genetic theory that they cultivated in the 1920s (Provine 1971). Figure 1 shows the impact of just three publications by these authors: Fisher (1930; 12,618 citations), Wright (1931; 5,493 citations), and Haldane (1932; 1,463 citations). Wright (1931) is a major paper rather than a book, but it is an important and timely summary of his views. These three works differ dramatically in style, just as the authors differed stunningly in personality. Although they contributed to this same synthetic period, members of this triumvirate held contrasting views about evolutionary process (Provine 1971, 1985, 1986; Frank 2012). Indeed, adherents to positions attributed to these authors continue to argue to the present time (Coyne et al. 1997, 2000; Wade and Goodnight 1998; Goodnight and Wade 2000).

The period 1937–1950 is sometimes known as the Modern Synthesis, a moniker that seems increasingly outdated. Figure 1 shows the impact of a major book by each of seven authors during this period: Dobzhansky (1937; 4,591 citations), Goldschmidt (1940; 1,009 citations), Mayr (1942; 4,380 citations), Huxley (1942; 1,891 citations), Simpson (1944; 1,684 citations), Schmalhausen (1949; 841 citations), and Stebbins (1950; 3,506 citations). Synthesis during this period is sometimes attributed more narrowly to Dobzhansky, Mayr, Simpson, and Stebbins, all of whom published books in a series edited by L. C. Dunn for Columbia University Press. Authors during this synthetic period did not always see eye to eye. For example, Mayr reports that he was so infuriated by Goldschmidt's book that he set to work on his own account of geographic variation and speciation (Mayr 1942, 1997). Mayr chastised Simpson for his brief treatment of speciation. Simpson responded that he did in fact devote considerable space to speciation (Simpson 1984, p. xx). The Modern Synthesis was far from harmonious.

Stepping back from the specifics of past synthetic periods, we see a continuous and accelerating graph of synthetic activity in figure 1. What about synthesis in our time? A striking feature of what I call the Ongoing Synthesis is that it extends beyond much-heralded advances in genomics (Losos et al. 2013) and covers the full range of evolutionary subdisciplines, including phenotypic evolution. Furthermore, many contemporary syntheses are comparable in integration, novelty, and scope to works of the Modern Synthesis. To continue comparing past and present periods of synthesis, I now turn to a comparison of Simpson's synthesis, and in particular his account of quantum evolution, with my own recent collaborative re-

search that bears on phenotypic evolution in deep evolutionary time.

Simpson (1944) and the Ongoing Synthesis in Evolutionary Quantitative Genetics

Simpson's 1944 book, *Tempo and Mode in Evolution*, is often cited as the first reconciliation of paleontology with Mendelian genetics. Simpson's own assessment was that he alone among the contributors to the Modern Synthesis steadfastly explored the time dimension and especially deep evolutionary time. Simpson's book strikes the modern eye as lacking in data tables, compilations, and graphs. In this sense, his argument is much closer to Darwin's than to, say, Schluter's (2000). Like Darwin, Simpson digests case studies out of sight and presents us with distilled principles. His use of theory is largely qualitative. He sketches arguments of logic and consistency using population genetic theory, cast in terms of the short-term dynamics of allelic fixation. To the modern reader, the disconnection between theory and observation is jarring, not only in *Tempo and Mode in Evolution* but in all the books of the Modern Synthesis. The problem is that population genetics is used as a theoretical framework, but nearly all observations in the 1930s and 1940s were in terms of continuously distributed (quantitative) traits, not in terms of allele frequencies. The disconnection was solved several decades later, principally by R. Lande (1988), who—building on the work of Pearson (1903), Dickerson (1955), Falconer (1960), Kimura (1965), Robertson (1966), and others—extended and used quantitative genetic theory to tackle evolutionary problems. Although the formal use of population genetic theory for continuously distributed traits is in the future, Simpson's theoretical arguments are often presented in graphical models that anticipate quantitative genetic models. The adaptive landscape concept is a case in point. Elsewhere I have elaborated on Simpson's role in adapting and applying Sewall Wright's adaptive landscape to the evolution of quantitative traits (Arnold et al. 2001), so here I will illustrate with just one key example.

Simpson's (1944) use of the adaptive landscape concept is explicit in his book and implicit in his vision of frequency curves for a quantitative trait, evolving within and transiting between adaptive zones (fig. 2). Simpson's concept of adaptive zones deserves some comment. By this term, Simpson means the morphological space that is occupied by an evolving lineage (or lineages) and that characterizes an ecological way of life. The connection of Simpson's model to contemporary evolutionary quantitative genetics becomes clear when we superimpose a modern stabilizing selection function, a function that could account for the short-term stability of the frequency curves

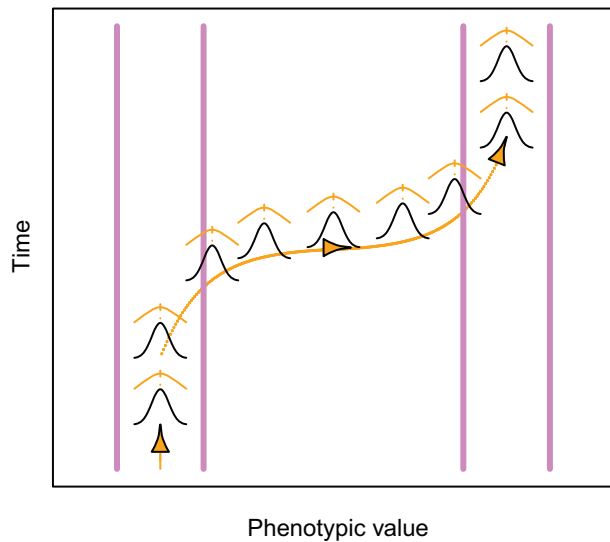


Figure 2: Simpson's (1944) concept of quantum evolution. Evolution is portrayed as translations in the positions of morphological traits, shown as (black) frequency curves. In Simpson's view, most evolution (phyletic evolution) occurred within adaptive zones (shown with purple boundaries), but on rare occasions a trait distribution made a rapid transition to a new adaptive zone (quantum evolution). Contemporary models for both kinds of evolution employ an adaptive landscape or fitness function (shown in orange), with an optimum that moves in morphological space and that is tracked by the trait mean.

that Simpson graphs. The optimum of that function anticipates the path of the evolving trait mean as it transits to a new adaptive zone (fig. 2), as in Lande's (1976) displaced optimum model. Simpson called rapid and substantial evolution of this kind "quantum evolution" and considered it his most controversial innovation in 1944. He used horse evolution as a possible example of quantum evolution, arguing that especially rapid and substantial evolution occurred as the horses moved from a browsing adaptive zone to a grazing adaptive zone. But Simpson also argued that such quantum evolution was exceptional and that most evolution occurred within adaptive zones. Such "phyletic evolution" was characterized by extinction and modest diversification within adaptive zones, as well as by literal stasis.

The Ongoing Synthesis in evolutionary quantitative genetics presents a vivid contrast with methodologies employed by Simpson and his colleagues during the Modern Synthesis. In the first place, quantitative genetics provides a theoretical framework with direct connections to data. Most traits involved in adaptation have continuous distributions and are affected by many genes. The frequency distributions of those traits are directly observable, unlike the frequencies of the many genes that contribute to phe-

notypic variation. Secondly, the key concepts in evolutionary quantitative genetics are rendered in statistical terms, with the consequence that they can be estimated in natural populations. Furthermore, key parameters of inheritance (e.g., heritability, h^2), selection (e.g., selection differentials and gradients), and population size (N_e) have been estimated in scores of populations of many dozens of taxa, especially over the past 30 years. Thirdly, those same key parameters play important roles in a flourishing family of models that can be used to characterize the statistical properties of an entire neutral or adaptive radiation. A useful property of these models is that maximum likelihood and related techniques can be used to compare the fits of models with alternative properties and to estimate the value of key parameters. Finally, huge data sets are increasingly available and are being used to test alternative models of phenotypic evolution. These data sets characterize the evolution of phenotypic traits such as body size in literally hundreds or thousands of lineages. For all of these reasons, evolutionary quantitative genetics provides a theoretical framework for achieving a new and more powerful synthesis than was possible in the period 1937–1950.

Two Examples of the Ongoing Synthesis

To summarize my argument to this point, in contemporary synthetic studies, we have a more useful conceptual framework, more powerful models and better ways of testing them, and much more data. Of course, these ingredients do not guarantee a better synthesis, but in fact we have good evidence that they are indeed having this anticipated effect. By way of illustration, I now take up the issue of how my colleagues and I have explored and tested Simpson's concept of quantum evolution. I will summarize just two studies (Estes and Arnold 2007 and Uyeda et al. 2011), but these are representative of a much larger contemporary effort with broader goals.

Analysis of the Gingerich Data

In Estes and Arnold (2007), we self-consciously applied five principles. In the process, we departed from a couple of unfortunate positions that had become entrenched in the literature. For example, instead of using generic models of process (e.g., unparameterized Ornstein-Uhlenbeck and Brownian motion models), we used models that were parameterized in terms of effective population size, inheritance, and selection. This improvement allowed us to test models at critical junctures by asking whether a model fit required parameter values that were biologically unrealistic. Second, we departed from the tradition of focusing on evolutionary rates. Let x be the amount of phenotypic

divergence during a time interval of length y . Much of the literature prior to Estes and Arnold (2007), graphed and analyzed evolutionary rates (x/y) as a function of time (y). This parameterization is fraught with a long-familiar ratio problem (Klauber 1939; Atchley 1978; Packard and Boardman 1988). In particular, the ratio is inevitably correlated with both x and y . One escapes from this problem by graphing and analyzing divergence (x) as a function of time (y).

Another aspect of our approach was to make use of published compilations of key parameters. Two examples are illustrated in figure 3. The heritability compilation reveals abundant genetic variation in most populations, especially for morphological traits, with a modal value of about 0.4. Calculations of a measure of stabilizing selection, ω^2 , and distance from the trait mean, \bar{z} , to an intermediate optimum, θ , are based on the data compiled by Kingsolver et al. (2001) and are described by Estes and Arnold (2007). Furthermore, a standardized measure of directional selection, β , shows a negative exponential distribution, with a modal value close to zero (Kingsolver et al. 2001; Kingsolver and Diamond 2011). In other words, most traits in most populations experience very weak directional selection, but occasionally strong values are recorded. Compilations of measures of stabilizing and disruptive selection reveal a pattern that is not completely understood. The distribution is unimodal, with the mode straddling the boundary between stabilizing and disruptive selection (Kingsolver et al. 2001; Stinchcombe et al. 2008; Kingsolver and Diamond 2011). If we focus on just the cases of stabilizing selection, the vision that emerges is one of trait means close to their optima, with small departures common and larger departures increasingly rare. We will want to keep these results in mind when we test models of evolutionary process that employ these same measures of inheritance and selection. When we conduct those tests and enter the dimension of deep evolutionary time, we do not want to abandon the insights we have gained from decades of microevolutionary study.

Our understanding of the connection between evolutionary process and pattern is being revolutionized by the use of stochastic models of evolutionary process. These models are central to the discussion that follows, so we need to understand them in some detail. Joseph Felsenstein, Russell Lande, and Michael Lynch are responsible for most of the nongeneric, parameterized models of phenotypic evolution that are used in Estes and Arnold (2007). That is to say, the models are parameterized using the measures of inheritance, selection, and population size that we just reviewed. The essence of these models can be grasped by imagining a set of replicate populations, all derived from the same ancestral population with a particular trait mean (call it zero) at generation zero. We

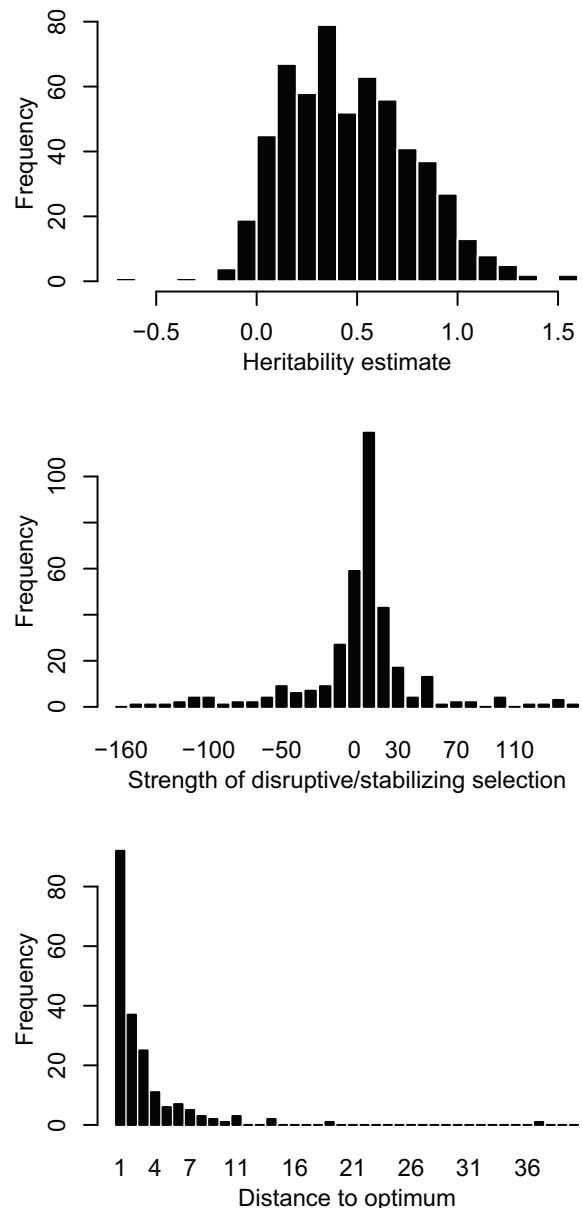


Figure 3: Distributions of key parameters of inheritance and selection as compiled from literature surveys over the past few decades (n is the number of estimates in each case). *Top*, Estimates of heritability for morphological traits ($n = 580$; Mousseau and Roff 1987; D. Roff, personal communication). *Middle*, Strength of stabilizing and disruptive selection, estimated as the width parameter (ω^2) of a Gaussian fitness function ($n = 355$). Positive values indicate downward curvature (stabilizing selection); negative values indicate upward curvature (disruptive selection; Kingsolver et al. 2001; Estes and Arnold 2007). *Bottom*, Distance from the trait mean to the optimum of the adaptive landscape in studies with stabilizing selection, measured in units of \sqrt{P} ($n = 197$; Kingsolver et al. 2001; Estes and Arnold 2007).

specify a set of stochastic rules for the evolution of that trait mean and suppose that all of the independent replicate lineages derived from the common ancestor obey those same rules. With the right specification, we can characterize the distribution of lineage means at any point in the future! The case of genetic drift supplies a concrete, simple example that is easy to follow. According to this model, the trait mean at a particular generation is the sum of two parts, the mean in the preceding generation and a deviation due to parental sampling. In other words,

$$\bar{z}(t) = \bar{z}(t-1) + N(0, \sigma^2), \quad (1)$$

where $\bar{z}(t)$ and $\bar{z}(t-1)$ are, respectively, the trait means in generations t and $t-1$. The second term on the right denotes a draw from a normal distribution with a mean of zero and a variance of $\sigma^2 = G/N_e$, where G is the additive genetic variance of the trait ($h^2 = G/P$) and N_e is effective population size (Lande 1976). The fact that the variance due to parental sampling is G/N_e is not a mystery. Suppose we take samples of size N_e from a population in which the genetic distribution of trait values is normal with a mean of zero and variance of G . The samples represent populations instantaneously derived from the ancestral population, and the variance among them will be G/N_e . Furthermore, and this is our key point, at any generation in the future, t , the population means will be normally distributed with a mean of zero and a variance of tG/N_e (Lande 1976). Results of this kind are the statistical equivalent of “replaying the tape of evolution,” to use Stephen Jay Gould’s (1989) metaphor. Of course, this statistical statement of future generations is much more powerful than the metaphor because we can use it as a foundation for testing our model with data.

We can use equation (1) to simulate a single lineage evolving by drift. The drifting mean undergoes a kind of random walk known as Brownian motion (fig. 4A). In this example, we have standardized within-population phenotypic variance at 1. Consequently, genetic variance equals heritability, which in this example is set at a constant, empirically established modal value of 0.4. For purposes of illustration, N_e is set at an unrealistically low value of 100. Simulation also allows us to portray the random walks of replicate lineages evolving by drift (fig. 4B). Because the lineage means are normally distributed at any given generation in this and in many other models, we can easily solve for and graph the 99% confidence limits for the model, shown here in blue. Everything we need to know about the model’s predictions is encapsulated in those confidence limits. For example, if we graph data on these same axes, we can do a quick visual model fit by overlaying the confidence limits on that data plot. See Estes and Arnold (2007) for examples.

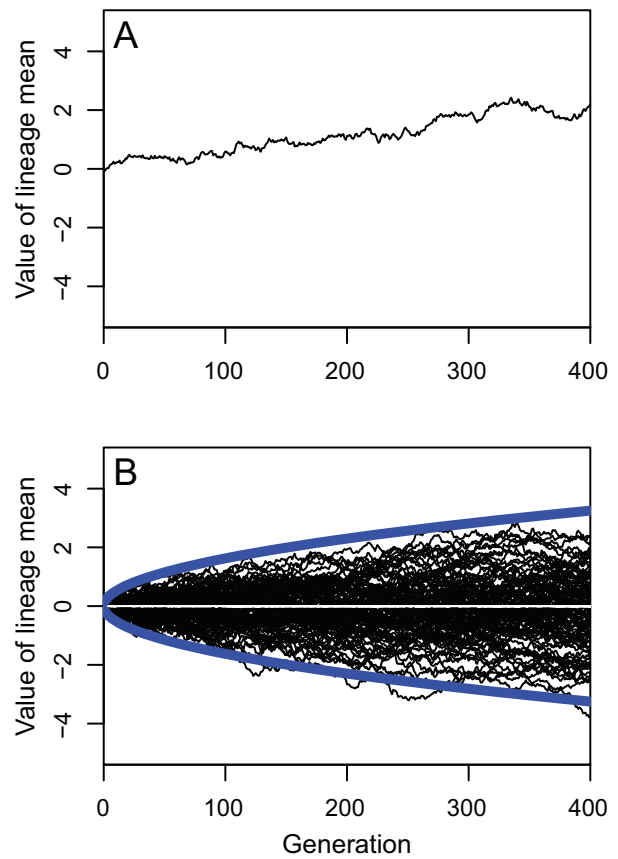


Figure 4: Simulations of trait means evolving by genetic drift. In all cases, the trait mean is zero at the starting point, heritability is 0.4, and effective population size is 100. *A*, A single simulation run representing the evolution of a single lineage. *B*, One hundred simulation runs, representing the evolution of 100 replicate lineages. The 99% confidence limits for the trait mean as a function of time are shown with thick blue lines.

Huge data sets that provide a statistical picture of the actual evolution of phenotypic traits have been compiled and made available over the past couple of decades. Philip Gingerich was one of the first to appreciate the importance of such compilations (Gingerich 1983, 1993, 2001). Although I will refer to the Gingerich data, much of the data on a microevolutionary timescale was assembled by Andrew Hendry and Michael Kinnison. Gingerich began compiling data from paleontological time series in the 1980s and combined his data with Hendry and Kinnison’s in Gingerich (2001), which may be consulted for details. Gingerich graciously made the data set available to us so that we could back-calculate divergence values from rates. The features of standardization in the Gingerich data mean that we can combine data from diverse studies in a meaningful way. The resulting, bulked picture of divergence tells

us what is often observed and what is virtually never observed in phenotypic evolution.

The Gingerich (2001) data are derived from time series in which a trait mean has been followed through time. The data are taken from 44 sources and consist of 2,639 values for change in trait mean over intervals ranging from 1 to 10 million generations. Some of the data are drawn from microevolutionary studies in which natural populations of living organisms have been sampled at two or more points in time, but most of the data represent studies of fossils, in which taxa have been sampled at different strata at the same site. The traits comprising the data are mainly size measurements representing the dimensions and shapes of shells and teeth, and some counts of various morphological structures. All of the trait means are standardized to a common scale of within-population, phenotypic standard deviation, \sqrt{P} . A great many animal taxa are represented, ranging from foraminiferans to ceratopsid dinosaurs.

Consider a single hypothetical data point, drawn from a particular time series (a set of estimated trait means, each separated by an estimated interval of elapsed time). Suppose the mean body size at generation 0 is 100 mm and at generation 100 is 150 mm. If the average within-population standard deviation in body size is 10 mm, then divergence is 150 mm – 100 mm = 50 mm, or $5\sqrt{P}$. The divergence interval is 100 = 10^2 generations. In a graphical display of this and similar data points, we will plot divergence ($5\sqrt{P}$) as a function of interval (10^2 generations). In the plots that follow, we add points, each time anchoring each point as if both divergence and time began at the origin. Note that divergence, as we have scaled it, is equivalent to a Felsenstein (1985) contrast, but our contrasts are not independent because some pairs of points share the measurement error associated with a common mean.

A plot of all the data reveals that stasis is the modal value for divergence, regardless of timescale (fig. 5A). Notice that the divergence values are clustered about zero in both the microevolutionary and fossil data sets. The strength of the clustering about zero divergence can be appreciated by focusing on the 99% confidence ellipse. The good fit of this very slightly positively inclined, narrow ellipse illustrates the remarkable fact that the overall regression of divergence on time is barely different from zero. In other words, the overall trend in the data is stasis, no change in mean, regardless of timescale. Furthermore, the fossil time series data appear to continue the trend of the microevolutionary time series data, a result that argues for continuity of process. The outer boundaries of this ellipse lie at about $6\sqrt{P}$ on either side of the divergence mean, which is close to zero. I have shown these boundaries in purple, because they will serve as a useful reference point in later plots. These boundaries harken back to

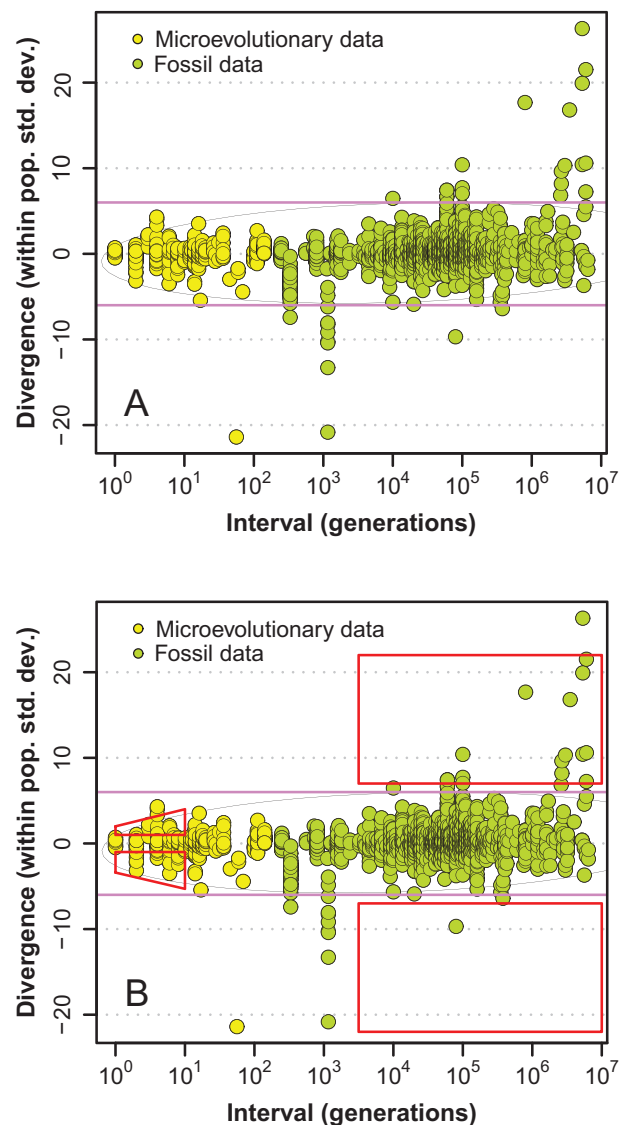


Figure 5: Gingerich (2001) data based on time series, showing trait divergence as a function of time. Divergence is measured in units of within-population phenotypic standard deviation, \sqrt{P} , and elapsed time is measured in generations and graphed on a log scale. The 99% confidence ellipse for the bivariate mean is defined with a thin gray line. The purple lines show divergence values of $6\sqrt{P}$ on either side of zero divergence. *A*, Data from microevolutionary (yellow) and fossil (green) studies show similar patterns. *B*, Models commonly failed to fit the data because of underprediction (no points predicted inside the small red quadrilaterals) or blowout (many points predicted inside the large red rectangles).

Simpson's concept of adaptive zones (fig. 2). Indeed, we can consider the purple boundaries in figure 5 as an estimate of the maximum value of width of adaptive zones.

Even though Estes and Arnold (2007) used a simple graphical approach in fitting models to the Gingerich data,

most models failed this screening test in spectacular fashion. The graphical screening procedure consisted of graphing the 95% confidence limits (CLs) for a particular model onto the 99% confidence ellipse for the data (Estes and Arnold 2007; fig. 6). In computing examples of the 95% CLs for each model, we used the entire range of possible values for key parameters (e.g., $h^2 = 0.0001$ – 0.9 , $\omega^2 = 3$ – 50 , $N_e = 10$ – $100,000$). Values outside these ranges are

improbable and may be biologically unrealistic. The models themselves represented the full range of explanation that has been offered for neutral and adaptive radiation over the past 35 years (Hansen and Martins 1996; Arnold et al. 2001). The upshot of implementing our procedure is that when a model spectacularly failed to account for the Gingerich data, it had been dealt a devastating blow.

Models failed in three distinct ways when Estes and

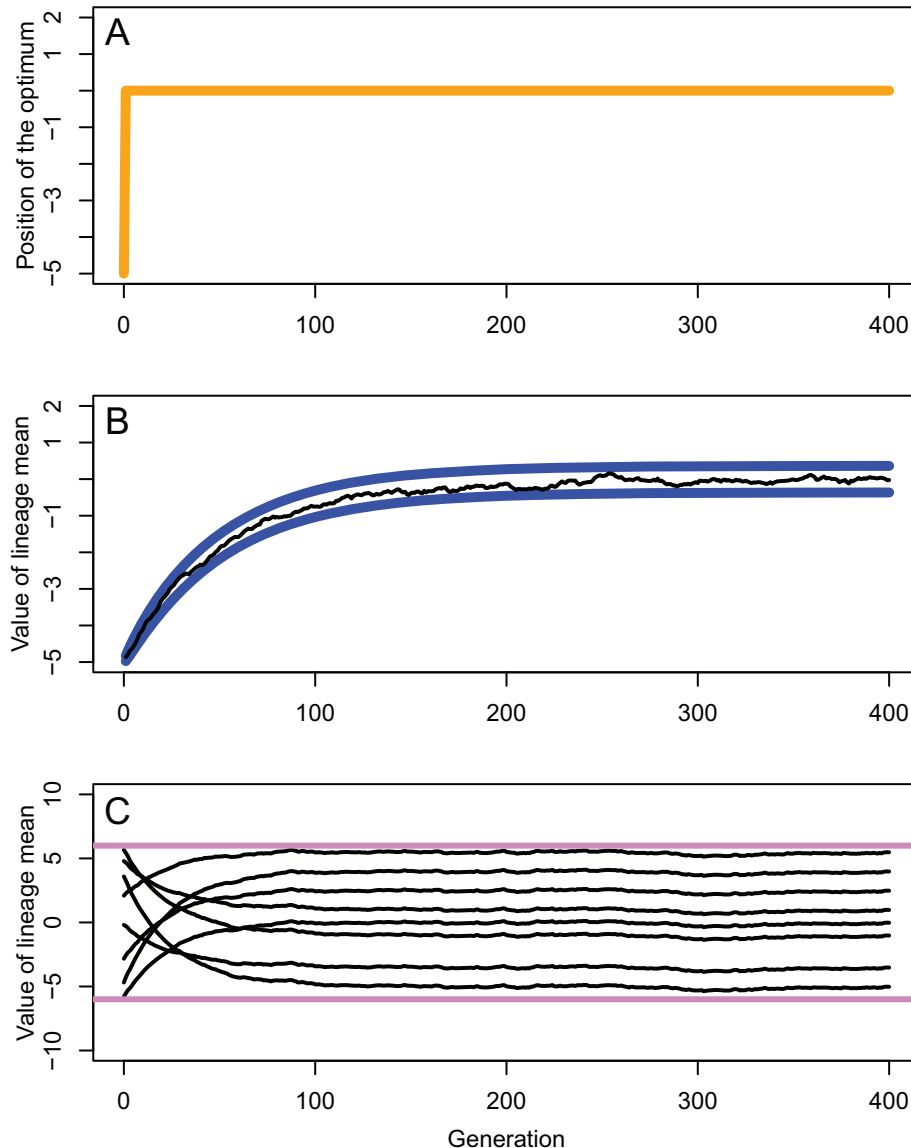


Figure 6: In Lande's (1976) displaced optimum model, an evolving trait mean tracks the movement of an intermediate optimum, θ . *A*, The optimum is displaced some distance from its initial position in a single generation. In this example, the distance is $5\sqrt{P}$. *B*, The trait mean rapidly evolves to a new position that approximates the new position of the optimum. The 99% confidence limits for the trait mean are shown as a function of time. Heritability is 0.4, and effective population size is 500. *C*, Lineage means evolving in response to optima that were displaced by differing amounts, all in the range of -6 to $+6\sqrt{P}$, at generation 1. Heritability and effective population size are as in *B*.

Arnold (2007) confronted them with data (table 1). In the first way (underprediction), the model fails to predict observed levels of divergence at short timescales (i.e., fails to account for the data inside the small red quadrilaterals in fig. 5B). For example, with realistic values of heritability and effective population size, models of genetic drift fail to predict enough divergence on short timescales. In the second, more common mode of failure (blowout), the model predicts too much divergence on long timescales (i.e., predicts an abundance of data inside the large red rectangles in fig. 5B). Again, genetic drift suffers from this problem. In the third kind of failure, the model does predict data inside the purple boundaries, but only if we invoke unrealistic values for key parameters. For example, peak shift models (in which the trait mean shifts between two adaptive peaks) require unrealistically small values for effective population size (<250), while a fluctuating optimum model with a single peak that moves by Brownian motion requires a value for stabilizing selection that is several orders of magnitude weaker than values commonly observed in nature.

When representatives of the entire family of existing stochastic process models confronted the Gingerich data, only a single model showed some promise of accounting for the data (table 1). Most of the models we tested failed to predict the Gingerich data for two or more of the three reasons just discussed. In nearly all cases, failure was dramatic rather than subtle. The basic problem for the models is that the data fall within a relatively narrow band of values (bounded by the purple lines in fig. 5), and most models blow out of those boundaries. In most cases, unrealistic parameter values had to be invoked to constrain blowout, but in some models blowout occurred under all combinations of parameter values (table 1). One model, however, the displaced optimum model proposed by

Lande (1976), could account for the data if just one of its parameters were varied. As we shall see when we assemble more data (Uyeda et al. 2011), the assertion that only the displaced optimum model can account for the data is a bit of an overstatement! In light of that additional data, we will need to resurrect the models with fluctuating optima (Brownian motion and white noise). Nevertheless, with only the Gingerich data in front of us, the displaced optimum model is the one that deserves our attention.

In Lande's (1976) displaced optimum model, the position of an intermediate optimum suddenly shifts to a new position in a single generation and thereafter resides at that position (fig. 6A). This model of peak movement represents a relatively sudden in situ change in environmental conditions or migration into a new environment. In the simulation shown in figure 6B, the lineage mean tracked the displaced optimum, a process that required about 100 generations. Thereafter, the mean fluctuated about the optimum in drift-stabilizing selection balance. The blue lines show the 99% confidence limits for the model.

By invoking displacements of the optimum that range from -6 to $+6$ within-population phenotypic standard deviations, we account for the full range of divergence within the purple boundaries but do not go outside them (fig. 6C). This success, however, has been achieved in the face of some obvious simplifications that are biologically unrealistic. First, we have contrived the sample of displacements to achieve the desired result. Second, all of the displacements occur in the first generation, and none happen thereafter. Recognizing these restrictions, we can easily imagine a model that relaxes them. For example, the displacements might occur at random according to some stochastic process (e.g., a Poisson process). Furthermore, the magnitude of the displacement once it occurs might

Table 1: Failure of models tested by Estes and Arnold (2007) to fit the Gingerich (2001) data due to underprediction, blowout, and unrealistic parameter values

Model	Underprediction	Blowout	Unrealistic parameter values
Genetic drift (Brownian motion)	Yes	Yes	Small h^2 , large N_e
Stationary optimum	Yes	No	Large ω^2 , large N_e
Fluctuating optimum (Brownian motion) ^a	Yes	Yes	Large σ_θ^2
Fluctuating optimum (white noise)	Yes	No	Large σ_θ^2
Steadily moving optimum (with white noise)	Yes	Yes	Small k
Peak shift (drift between two optima)	Yes	No	Small N_e
Displaced optimum	No	No	None

Note: Failure to fit the data is denoted with a "Yes." In the model with a steadily moving optimum, k is the deterministic, per-generation change in the position of the optimum, θ . In models with a fluctuating optimum, the position of that optimum varies, all or in part, because of departures represented by draws from a normal distribution with zero mean and a variance of σ_θ^2 .

^a Estes and Arnold (2007) gave an erroneous expression for the variance among lineage means for this model, which nevertheless provides a reasonable approximation on long timescales. See Hansen et al. (2008) for the correct expression.

be drawn from some probability distribution (e.g., a normal distribution centered about zero). We will explore both of these avenues of improvement in the next section. Before we leave the displaced optimum model, we should recognize one additional problem with its candidacy as an explanatory model; it cannot easily account for phylogenetic resemblance between sister taxa (Hansen 2012).

Analysis of the Uyeda Data

To continue our comparison of Simpson's (1944) synthesis with contemporary efforts, I now turn to the study by Uyeda et al. (2011). I will first discuss the data and then the models, before drawing conclusions from the model fits. Our first goal in this study was to improve the data set in a couple of ways. First, we restricted the focus to size-related traits so that we might realistically strive for a homogeneous process. Second, we more than doubled the amount of the data. Most of the increase in sample size was achieved by adding data on body size from time-calibrated evolutionary trees. Early in our collaboration, Hansen noted that although the Gingerich data support an overall model of stasis, comparative studies often show that sister taxa have similar body sizes. Such resemblance between sister taxa rapidly decays under a model of stabilizing selection (Hansen and Martins 1996), so either stabilizing selection is only part of the story or something is missing from the Gingerich data. With this apparent contradiction in mind, we decided to add tree-based data to our study. The upshot is a data set unprecedented both in size (over 8,000 divergence values from 206 studies) and temporal span (divergence intervals range from 1 year to more than 300 million years). Third, we focused primarily on vertebrates, again in the spirit of improving the prospects for a data set that might be modeled with a single, homogeneous process. For simplicity, I shall refer to these data as the Uyeda data even though they build on the Gingerich data set and were compiled as part of collaboration.

Why, in the plots that follow, do we use different axis scales than the ones employed by Estes and Arnold (2007)? In the first place, by adding tree-based data, we greatly expanded the magnitude of divergence that we observed. To accommodate this expansion, we used a log scale that is equivalent to the proportional change in body size. Using our previous hypothetical data point (100 mm at generation 0, 150 mm at generation 100), the proportional change in body size is $(150 - 100)/100 = 0.5$, or a 50% increase in body size. The second change occurs on the X-axis. Before, we measured time in generations, now we measure it in years. Although the same patterns are evident on both timescales, the pattern appears blurred on the generational scale, perhaps because of error in the esti-

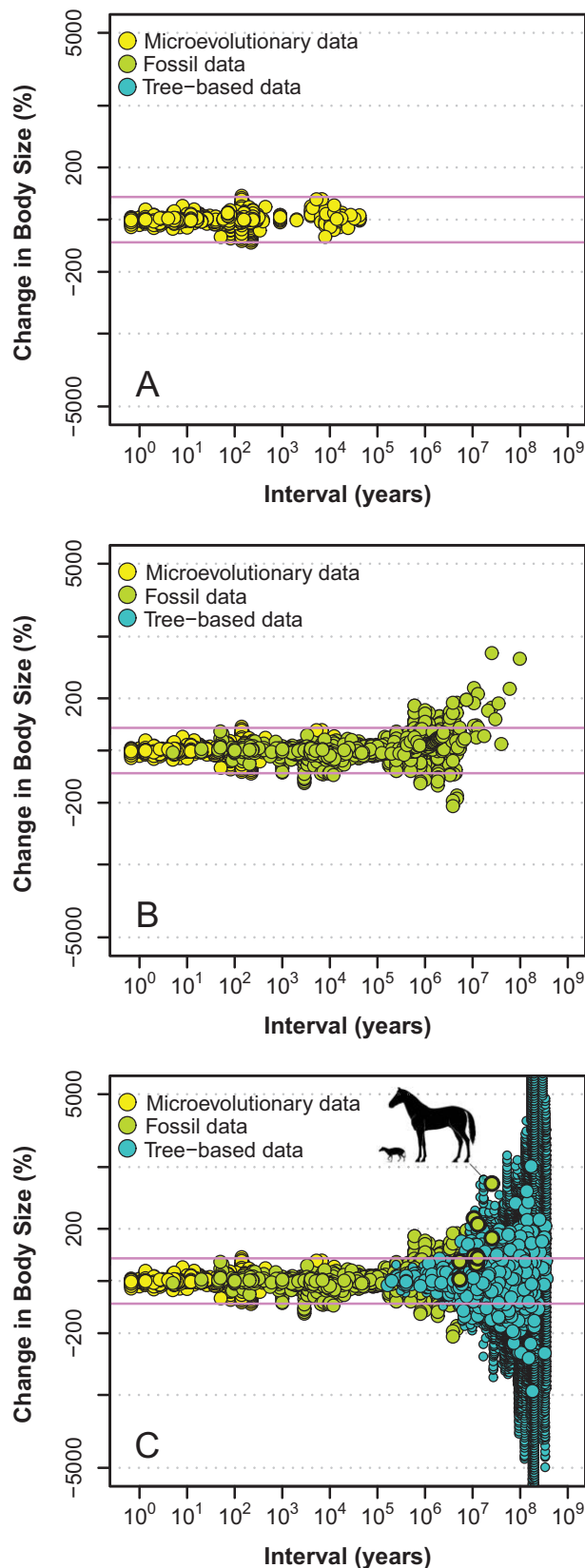
mates of generation time (see Uyeda et al. 2011 for comparison plots).

In figure 7, we add data to the plot in the order: microevolutionary, fossil, and, finally, tree-based. Notice that the pattern merges smoothly as the data types are added, again suggesting homogeneity of pattern and perhaps of process. However, with the addition of fossil time series data in this plot (fig. 7B), we see an inkling that at long timescales the pattern is a different one from that seen in the microevolutionary data. In particular, we see an abundance of fossil data in which divergence exceeds a 100% change in body size. Adding in the tree-based data (fig. 7C), we see many more examples in which divergence in body size exceeds 200%. Two kinds of tree-based data points are shown in blue. Note that when data (contrasts) are drawn from a tree, they have no sign. For the purposes of merging the tree-based data with the microevolutionary and fossil data (which have signs), we assigned signs at random to the tree-based contrasts. The larger blue circles show values of divergence that are averaged about a node on a tree. The smaller blue circles show divergence values for individual pairs of divergence values on a tree.

We can gain perspective on the pattern before us if we ask for the position of the purple boundary lines that we drew on the Gingerich data (fig. 5). The equivalent boundaries for this divergence scale, also shown in purple, are $\pm 65\%$ change in body size (fig. 7C). In other words, the vast majority of size evolution in vertebrates appears to be limited to a $\pm 65\%$ change in body size, regardless of timescale. Divergence in body size greater than 100% is common only on timescales greater than 1 million years.

Another perspective on the data is gained by comparing the pattern to the barrel of a blunderbuss, a muzzle-loading firearm with a flared barrel opening that was used in the seventeenth and eighteenth centuries. On a timescale of less than 1 million years, size divergence is restricted to a narrow range that resembles the slender base of the barrel. On longer timescales, divergence values flare out much like the flared end of the barrel. Because these divergence values are well outside the purple lines denoting bounded evolution, it is tempting to view them as examples of Simpson's quantum evolution. Indeed, the example of size divergence from *Ephippus* to modern horses (fig. 7) strengthens this view. More generally, How can we account for the blunderbuss pattern? This question will be our main concern in the sections that follow.

The revelation that very substantial divergence can occur on very long timescales (i.e., the flared end of the blunderbuss) makes us take another look at some of the models we discarded in the Estes and Arnold (2007) study. In particular, Brownian motion models are capable of producing the blunderbuss pattern (fig. 6A, 6C of Estes and Arnold 2007). These models are of two types: Brownian



motion of a trait mean (fig. 4) and Brownian motion of an intermediate optimum. Furthermore, white noise motion of the optimum could account for the long, slender base of the blunderbuss barrel, even though the end of the barrel would not flare dramatically as in the blunderbuss (fig. 6B of Estes and Arnold 2007). In other words, the new information about pattern, supplied by the tree-based data, causes us to resuscitate two models prematurely pronounced dead by Estes and Arnold (2007), Brownian motion and white noise motion of the optimum, as well as to explore relaxed versions of the displaced optimum model. We need a strategy to efficiently explore all of these model candidates.

While the model comparisons of Estes and Arnold (2007) do not account for the flared end of the blunderbuss, their conclusions about the long barrel of the blunderbuss remain valid. With that perspective in mind, the Uyeda et al. collaborators decided to focus on models that might account for the flared end of the blunderbuss and to let a surrogate, place-keeping process account for the long, slender base of the barrel. We used white noise fluctuation of the lineage mean about trait optimum as that surrogate process. This process is simpler than the drift process we simulated in figure 4. Under a white noise process, the value of the trait mean of each generation is simply a draw from a normal distribution with zero mean and a variance of σ_p^2 , whereas under drift the deviation of the mean from its value in the preceding generation is normally distributed; that is, $\bar{z}(t-1)$ is dropped from the right side of equation (1). We employed a white noise process in full knowledge of the fact that tracking an optimum that fluctuates across the full width of the central band of data will inevitably place an evolving lineage in hazardous demographic territory. As the distance from the trait mean to the optimum increases, mean fitness in the population declines and the risk extinction increases. The compensating advantage of a white noise model is that it requires a single variance term, and it can be easily com-

Figure 7: Uyeda et al. (2011) data drawn from time series and time-calibrated trees, showing trait divergence as a function of time and revealing a blunderbuss pattern. Divergence is measured as proportional change in the mean and graphed on a log scale. Elapsed time is measured in years and graphed on a log scale. *A*, Microevolutionary data only. *B*, Microevolutionary and fossil data. *C*, Combined plot of the microevolutionary, fossil, and tree-based data, revealing the blunderbuss pattern. The larger blue circles show values of divergence that are averaged about a node on a tree. The smaller blue circles show divergence values for individual pairs of divergence values on a tree. Fossil data points representing trait divergence during the quantum evolution of horses are outlined in black and superimposed on the tree-based data. The two purple lines denote $\pm 65\%$ change in size-related traits, equivalent on these axes to the purple lines shown in figure 5.

bined with more complicated models that might account for the flared end of the barrel. In particular, we used white noise (WN) alone, as well as WN combined with Brownian motion and WN combined with two descendants of the displaced optimum model (the single- and multiple-burst models), which we will now describe. These two models also harken back to a predecessor model, which was not evaluated by Estes and Arnold (2007). In that Hansen (1997) model, a stabilizing selection optimum maintains a long-term position but also experiences rare shifts in position.

Two key features of the single-burst model are illustrated in figure 8A. The first key feature is that the optimum can move away from its initial position at zero in a single burst. The timing of this single burst is dictated by a Poisson process. The simulations illustrated in figure 8 used an artificially large value of the Poisson parameter λ so that bursts happen in just a few hundred generations. Once that single burst occurs, the optimum retains its new position ever after. The second key feature of the model is that magnitude of the burst (the displacement of the op-

timum) is a draw from a normal distribution with a mean of zero and a variance of σ_D^2 . For example, in figure 8A we see that the single burst occurs at about generation 20 and amounts to a displacement of about $8\sqrt{P}$ away from the starting position. A displacement of this magnitude carries the long-term lineage mean outside the $\pm 6\sqrt{P}$ boundary lines shown in purple. If we parameterize the model so that σ_D^2 is smaller, single bursts will fall within the limits of the adaptive zone so that we have a model for within-adaptive zone, phyletic evolution.

The obvious point that an evolving lineage can experience multiple bursts is illustrated in figure 8B. In the illustrated implementation of a multiple-burst model, five bursts occurred, and two of those, at about generations 500 and 750, carried the lineage mean outside the purple boundary lines. As in the single-burst model, the timing of bursts is dictated by a Poisson process, and the magnitudes of the bursts are draws from a normal distribution with zero mean. To determine the best choice of parameters for the Poisson process and the normal distribution in each type of model, we used maximum likelihood.

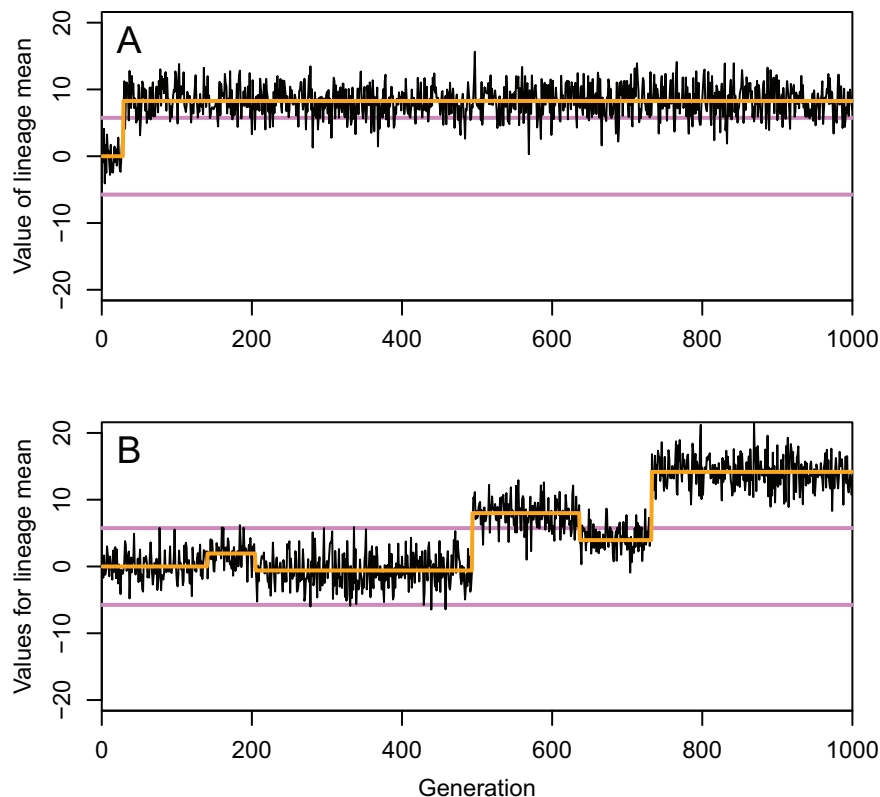


Figure 8: Simulations of the single- (A) and multiple-burst (B) models showing evolving trait means as a function of time. The position of the intermediate optimum, θ , is shown in orange; the position of the evolving trait mean is shown in black. In these simplified models, the trait mean shows white noise fluctuation about the optimum and instantaneously tracks any movement of the optimum ($\sigma_p^2 = 5$, $\sigma_D^2 = 37$, $\lambda = 0.005$). Purple lines are as in figure 5.

The fits of the four models to the Uyeda data are illustrated in figure 9 and compared in table 2. For each model, we can characterize the expected distribution of data at any given point in time. The first three models yield relatively simple normal distributions, but in the fourth, the distribution is more complicated. Using those expressions, we can calculate the likelihood of the data for any given choice of parameter values. For each model, we proceed by finding the choice of parameter values that maximizes the likelihood of the data. In the middle column of table 2, we show the maximum likelihood estimates for the white noise parameter, expressed as a standard deviation. Notice that for all three of the combined models, the estimates of this white noise standard deviation are very similar, a point that we will revisit shortly. The column on the far right shows the Akaike Information Criterion (AIC) values, which represent best fits for each model, corrected for differences in numbers of parameters. The asterisk on the AIC label reminds us that we have treated the data points as independent observations in calculating AIC (i.e., we have ignored the covariance structure of the data), so our conclusions in comparing the models are necessarily tentative. The pattern of these best fits are also illuminating (fig. 9). The blue lines in figure 9 show the 95% CLs for the three best-fitting models. Although the best-fit CLs for the white noise model are not shown, they are easily visualized as two parallel lines, slightly further apart than the three sets of lines describing the blunderbuss barrel in figure 9 but without the flaring on the far right. We see from the AIC column (table 2) that the white noise-only model appears to fare appreciably worse than the other models. The AIC also suggests that the multiple-burst model fits best, followed by single-burst and then by Brownian motion. This order seems to be contradicted

by the plots of best CLs. In particular, the Brownian motion fit appears to be as good as the multiple-burst fit. The Brownian motion model predicts a normal distribution of data at any given point in time, but a close examination of the data reveals a nonnormal overabundance of divergence data near the zero point, especially at the far-right end of the plot. The single- and multiple-burst models do a better job of accounting for this aspect of the data distribution, and those better jobs are reflected in their AIC scores.

At this point, I would like to retract a conclusion that I reached with Suzanne Estes in Estes and Arnold (2007). In that paper, we concluded that both versions of the Brownian motion model could be discarded because both models predicted more and greater values of divergence on long timescales than we observed in the Gingerich data. On that basis, we argued that Felsenstein's (1985) independent contrasts foundation for comparative studies, which relies on a Brownian motion model of evolution, might be built on sand. However, comparative studies supply the very values of large divergence that Brownian motion predicts. Consequently, Felsenstein (1985) is correct in arguing that Brownian motion might be a good approximation to phenotypic evolution. We must, however, add a caveat. Brownian motion of the adaptive peak may be a good approximation, but Brownian motion of the trait mean about a stationary peak (i.e., genetic drift) is not. We can reach this conclusion by asking how large effective population size must be to account for the blunderbuss of CLs shown in figure 9 on the far right. The answer is hundreds of millions of individuals, if heritability takes the empirically observed modal value of 0.4. The problem for the genetic drift version of Brownian motion is that unrealistically large population size is required to

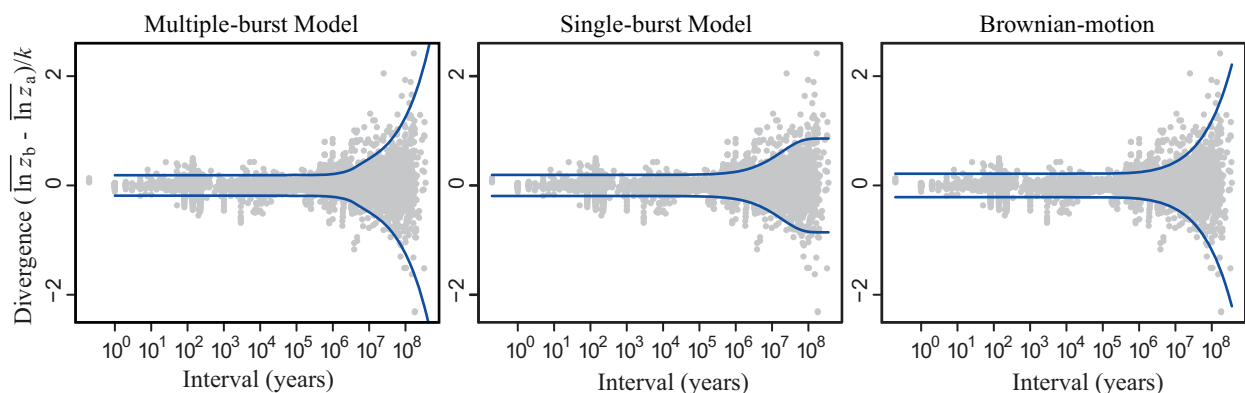


Figure 9: Graphical comparison of the best-fit versions of the multiple-burst, single-burst, and Brownian motion models. Silhouettes of the data points are shown in gray (same data shown in fig. 7C, except that pairwise tree data are not included). The blue lines show the 99% confidence limits corresponding to the parameters fit by maximum likelihood. Divergence is measured as the difference between the means of log-transformed size, standardized by dimensionality.

Table 2: Comparison of model fits to the Uyeda et al. (2011) data

Model	ML estimate of white noise parameter, σ_p	AIC*
White noise only	.20	-2,941
Brownian motion + white noise	.11	-7,878
Single-burst + white noise	.10	-9,018
Multiple-burst + white noise	.10	-9,143

Note: The maximum likelihood (ML) estimates for the white noise parameter are expressed as a standard deviation. The test statistic (AIC*) does not account for the covariance structure of the data. The asterisk indicates that we have treated the data points as independent observations in calculating Akaike Information Criterion (AIC; i.e., we have ignored the covariance structure of the data), so our conclusions in comparing the models are necessarily tentative.

stop the lineage from drifting outside the bounds of the slender barrel of the blunderbuss and delaying substantial divergence until tens or hundreds of millions of years have elapsed.

What can we infer about quantum evolution from the maximum likelihood estimates of parameters for the best-fitting multiple-burst model? Consider first the normal distribution of white noise deviations from zero on the *Y*-axis, the initial position of the optimum (fig. 10). The best estimate of the standard deviation of this distribution (shown with a dashed line at the upper left) is about 0.10 (table 2). In contrast, the best-fit normal distribution of burst sizes (shown with a solid line) has a standard deviation of 0.27, about three times larger (right-hand panel in fig. 10). Turning finally to the best-fit distribution of burst waiting times (bottom panel in fig. 10), we see that on a log scale the distribution is skewed toward “short” waiting times, but the mean time is 25 million years and bursts are very uncommon in intervals shorter than 1 million years. In other words, evolutionary bursts that might carry a lineage out of an adaptive zone are very rare. Furthermore, the large divergence values observed in some fossil time series and more commonly in comparative data would require a succession of several to many bursts.

Landis et al. (2013) have described a model that combines Brownian motion with evolutionary bursts. One can interpret this model as Brownian motion of an optimum that also experiences rare bursts of movement, described by a Poisson process. Fitting this model to tree-based primate data on brain size revealed a million-year waiting time for bursts, similar to the waiting time described by Uyeda et al. (2011).

Conclusions and Perspectives

What Is the Best Model for Phenotypic Evolution?

In this section, I summarize and discuss essential results while maintaining my focus on connections to key con-

cepts pioneered by G. L. Simpson (adaptive zones, phyletic evolution, quantum evolution). In the process, I neglect many worthy and contentious topics. The issue of how stabilizing selection might be maintained on a geological timescale, for example, is discussed at length by Estes and Arnold (2007).

Micro- and meso-evolution (Simpson’s phyletic evolution) are bounded so that diversification occurs within adaptive zones that have distinct limits or bounds. Bounded diversification can be visualized as the long barrel of a blunderbuss. The bounds define a region of trait divergence only 6 standard deviations on either side of the long-range trait mean. For size-related traits, this region represents change of less than 65% in trait mean. Remarkably, most evolution occurs within these boundaries on all timescales.

What is the explanation for this bounded evolution? Although the existence of bounds on a geological timescale seems unequivocal, the process that produces and maintains evolutionary bounds is a topic of current debate. The bounded evolution revealed by the slender barrel of the blunderbuss should not be confused with stasis. The limits of that slender barrel lie ± 6 within-population standard deviations, or a $\pm 65\%$ change in body size, on either side of literal stasis. Within those bounds, individual species can appreciably evolve. Indeed, though circumscribed, those bounds are wide enough to accommodate the differentiation of many (most?) vertebrate genera and families. For this reason, the barrel of the blunderbuss is comparable to an adaptive zone in Simpson’s (1944) sense, that is, bounds within which considerable but constrained phyletic evolution can take place. In other words, the barrel of the blunderbuss is not a narrow band of literal or even modest stasis. For this reason, many processes and hypotheses that have been put forth to explain long-term stasis do not provide a satisfactory explanation for our slender barrel. Take, for example, the idea that gene flow and complex geographic structure within species prevents range-wide incorporation of novel excursions in average

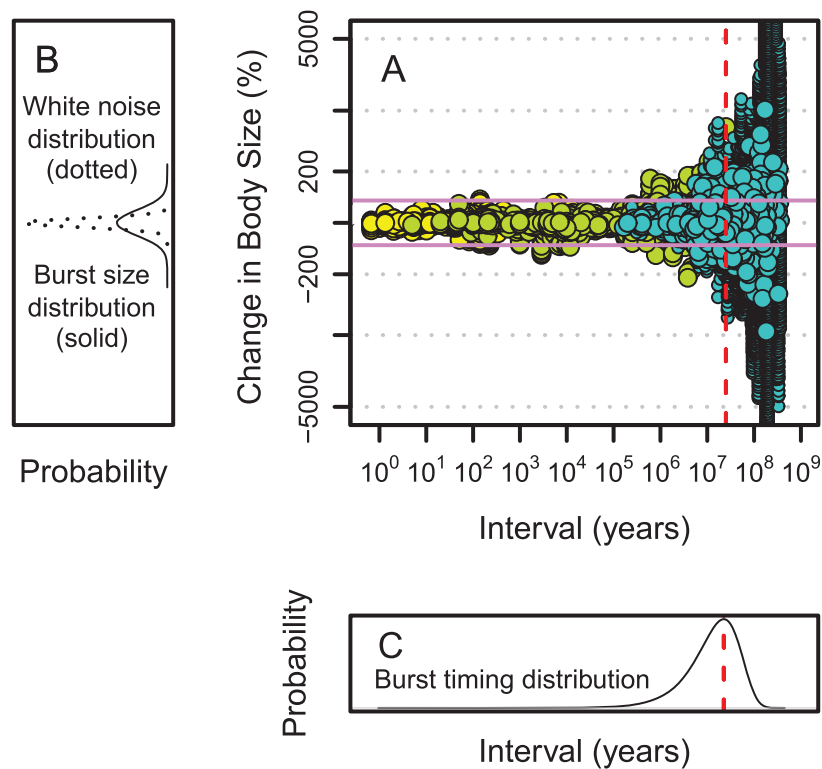


Figure 10: Maximum likelihood estimates of parameters for the best-fitting multiple-burst model. A, Plot of the Uyeda et al. (2011) data, as in figure 7. Gaussian curves representing the white noise distribution (mean = 0, standard deviation = 0.096) and the burst size distribution (mean = 0, standard deviation = 0.272); these standard deviation values are for the vertical scale shown in figure 9; other conventions are as in figure 7. C, Burst timing distribution, showing the waiting times between bursts ($1/\lambda$) obtained by bootstrapping over studies (2,000 replicates). The vertical red dashed lines show the position of the average waiting time on the actual and simulated data plots.

phenotype (Futuyma 1987; Eldredge et al. 2005). Hybridization might tend to pull the range-wide species mean back toward a central value, but it cannot explain why independently evolving species and genera remain inside the slender barrel. In other words, complex geographic structure might help explain stasis within species but does not help us account for long-maintained adaptive zones. The explanation is likely to lie in the realm of community ecology (Futuyma 2010), far outside the domain of genetic constraints and population structure.

Rare bursts of evolution can carry trait means into new adaptive zones, beyond the bounds of normal phyletic evolution. These rare events can be modeled as draws of adaptive peak movement from a statistical distribution with a mean of zero. Most evolutionary bursts are modest in magnitude and do not carry a lineage outside its adaptive zone. Irrespective of magnitude, bursts are rare with an average waiting time of 25 million years. Despite this rarity, on timescales of 100 million years or more, bursts are increasingly inevitable. A succession of these stochastic but inevitable bursts can carry a lineage far beyond the

boundaries of its original adaptive zone (Simpson's quantum evolution). The flared end of the blunderbuss reflects the accumulation of these rare bursts of evolution.

It is tempting but misleading to equate the multiple-burst and related models with punctuated equilibrium (Eldredge and Gould 1972). The problem with punctuated equilibrium is that it is an idea that has different implications to different people. It is not a well-defined model. Consequently, punctuated equilibrium does not provide a useful touchstone for current modeling efforts. Pennell et al. (2013) have highlighted this problem by pointing out that punctuated equilibrium conflates four separate questions: whether evolution is gradual or pulsed, whether evolution is concentrated at speciation or within lineages, whether change at speciation is adaptive or neutral, and whether species selection shapes among-lineage diversity. To make matters worse, some of these issues appear unanswerable. In light of these problems, Pennell et al. (2013) conclude that punctuated equilibrium has outlived its usefulness. I agree.

We will need additional studies on additional kinds of

traits and other kinds of organisms to determine whether the blunderbuss pattern, documented for size-related traits in vertebrates, is a general pattern. One interpretation of evolutionary bursts is that they coincide with the invasion of new adaptive zones, but other interpretations are possible and need to be tested.

*Some General Lessons about Synthesis
in Evolutionary Biology*

Synthesis is not something that we should anxiously await. Synthesis has been an ongoing activity since 1963 and appears to be accelerating.

We should not expect harmony during periods of synthesis. During long-recognized periods of great synthesis in evolutionary biology (e.g., 1930–1932, 1937–1950), participants often disagreed, sometimes strenuously, about minor and major topics. Likewise, in our own time, dissension and discord are all about us.

To synthesize, we need diverse perspectives and bridges between them. Uyeda et al. (2011), for example, discovered the blunderbuss pattern by combining data from three types of studies (microevolutionary, paleontological, and comparative).

My colleagues and I pioneered a new approach for evaluating and testing stochastic models of evolution by breaking with tradition and plotting the data so that they communicated directly with the models. The general point is that theory needs to be tested with the best available data. We also need to use the most powerful versions of theory as we strive to move beyond metaphors.

Are we in the midst of one of the great synthetic periods in evolutionary biology? I have argued that we are in such a period because of an extraordinary ongoing surge in citations of works that span the full range of topics in our field. We should not be distracted by the attention lavished on a few hot areas. The Ongoing Synthesis is broad and diverse.

The Ongoing Synthesis is comparable in its breadth and integration to earlier periods of synthesis. I have argued that two recent studies in which I participated extend Simpson's (1944) synthesis in a significant way and are emblematic of a new synthetic wave. In particular, the use of large data sets to test alternative models of evolutionary process is moving us toward new and powerful generalizations about phenotypic evolution. Examples of this new approach include historical tests for alternative selection regimes (Butler and King 2004; Hansen et al. 2008), inferring rates of clade diversification on phylogenetic trees (Alfaro et al. 2009; Meredith et al. 2011; Stadler 2011), and tests for alternative models of trait evolution, along the lines of the ones described here (Hunt 2007; Harmon et al. 2010).

Acknowledgments

I am grateful to my colleagues S. Estes, T. Hansen, J. Pienaar, and J. Uyeda for the opportunity to collaborate on the two projects that are summarized here. I am grateful to D. Futuyma, T. Hansen, and M. McPeck for critically reading the manuscript. Thanks to J. Uyeda for help with figures, to P. Gingerich and D. Roff for providing raw data, and to R. Bürger, A. Hendry, A. Jones, R. Lande, and D. Maddison for helpful discussions. J. Felsenstein graciously provided results for the Brownian motion model of peak movement that identified an error in expression (11) of Estes and Arnold (2007). Thanks especially to the many investigators whose data were compiled and analyzed. The preparation of this article was supported by an OPUS grant from the National Science Foundation (DEB-0947162).

Literature Cited

- Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carneval, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the USA* 32:13410–13414.
- Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* 112/113:9–32.
- Atchley, W. R. 1978. Ratios, regression intercepts, and the scaling of data. *Systematic Zoology* 27:78–83.
- Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist* 164:683–695.
- Coyne, J. A., N. H. Barton, and M. Turelli. 1997. Perspective: a critique of Sewall Wright's shifting balance theory of evolution. *Evolution* 51:643–671.
- . 2000. Is Wright's shifting balance process important in evolution? *Evolution* 54:306–317.
- Dickerson, G. E. 1955. Genetic slippage in response to selection for multiple objectives. *Cold Spring Harbor Symposium on Quantitative Biology* 20:213–224.
- Dobzhansky, T. 1937. *Genetics and the origin of species*. Columbia University Press, New York.
- Eldredge, N., and S. J. Gould. 1972. Punctuated equilibrium: an alternative to phyletic gradualism. Pages 82–115 *in* T. Schopf, ed. *Models in paleobiology*. Freeman, Cooper, San Francisco.
- Eldredge, N., J. N. Thompson, P. M. Brakefield, S. Gavrillets, D. Jablonski, J. B. C. Jackson, R. E. Lenski, B. S. Lieberman, M. A. McPeck, and W. Miller III. 2005. The dynamics of evolutionary stasis. *Paleobiology* 31:133–145.
- 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.
- Falconer, D. S. 1960. *Introduction to quantitative genetics*. Ronald, New York.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.

- Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon, Oxford.
- Frank, S. A. 2012. Wright's adaptive landscape versus Fisher's fundamental theorem. Pages 41–57 *in* E. Svensson and R. Calsbeek, eds. The adaptive landscape in evolutionary biology. Oxford University Press, New York.
- Futuyma, D. J. 1979. Evolutionary biology. Sinauer, Sunderland, MA.
- . 1987. On the role of species in anagenesis. *American Naturalist* 130:465–473.
- . 2010. Evolutionary constraint and ecological consequences. *Evolution* 64:1865–1884.
- . 2011. Expand or revise? the evolutionary synthesis today. *Quarterly Review of Biology* 86:203–208.
- Gingerich, P. D. 1983. Rates of evolution: effects of time and temporal scaling. *Science* 222:159–161.
- . 1993. Quantification and comparison of evolutionary rates. *American Journal of Science* 293:453–478.
- . 2001. Rates of evolution on the time scale of the evolutionary process. *Genetica* 112/113:127–144.
- Goldschmidt, R. 1940. The material basis of evolution. Yale University Press, New Haven, CT.
- Goodnight, C. J., and M. J. Wade. 2000. The ongoing synthesis: a reply to Coyne, Barton, and Turelli. *Evolution* 54:317–324.
- Gould, S. J. 1980. Is a new and general theory of evolution emerging? *Paleobiology* 6:119–130.
- . 1989. Wonderful life: the Burgess Shale and the nature of history. Norton, New York.
- Haldane, J. B. S. 1932. The causes of evolution. Longmans, Green, London.
- Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351.
- . 2012. Adaptive landscapes and macroevolutionary dynamics. Pages 205–226 *in* E. I. Svensson and R. Calsbeek, eds. The adaptive landscape in evolutionary biology. Oxford University Press, New York.
- Hansen, T. F., and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–1417.
- Hansen, T. F., J. Pienaar, and S. H. Orzack. 2008. A comparative method for studying adaptation to a randomly evolving environment. *Evolution* 62:1965–1977.
- Harmon, L. J., J. B. Losos, J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. Kozak, et al. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385–2396.
- Hunt, G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. *Proceedings of the National Academy of Sciences of the USA* 104:18404–18408.
- Huxley, J. 1942. Evolution: the modern synthesis. Allen & Unwin, London.
- Kimura, M. 1965. A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences of the USA* 54:731–736.
- Kingsolver, J. G., and S. E. Diamond. 2011. Phenotypic selection in natural populations: what limits directional selection? *American Naturalist* 177:346–357.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. *American Naturalist* 157:245–261.
- Klauber, L. M. 1939. A statistical study of the rattlesnakes. VI. Fangs. *Occasional Papers of the San Diego Society of Natural History* 5:3–61.
- Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30:314–334.
- . 1988. Quantitative genetics and evolutionary theory. Pages 71–84 *in* B. Weir, E. Eisen, M. Goodman, and G. Namkoong, eds. *Proceedings of the Second International Conference on Quantitative Genetics*. Sinauer, Sunderland, MA.
- Landis, M. J., J. G. Schraiber, and M. Liang. 2013. Phylogenetic analysis using Levy processes: finding jumps in the evolution of continuous traits. *Systematic Biology* 62:193–204.
- Losos, J. B., S. J. Arnold, G. Bejerano, E. D. Brodie III, D. Hibbett, H. E. Hoekstra, D. P. Mindell, et al. 2013. Evolutionary biology for the 21st century. *PLoS Biology* 11:e1001466, doi:10.1371/journal.pbio.1001466.
- Mayr, E. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. Columbia University Press, New York.
- . 1963. Animal species and evolution. Harvard University Press, Cambridge, MA.
- . 1997. Goldschmidt and the evolutionary synthesis: a response. *Journal of the History of Biology* 30:31–33.
- Meredith, R. W., J. E. Janečka, J. Gatesy, O. A. Ryder, C. A. Fisher, E. C. Teeling, A. Goodbla, et al. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* 334:521–524.
- Mousseau, T. A., and D. A. Roff. 1987. Natural selection and the heritability of fitness components. *Heredity* 59:181–197.
- Packard, G. C., and T. J. Boardman. 1988. The misuse of ratios, indices, and percentages in ecophysiological research. *Physiological Zoology* 61:1–9.
- Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 200:1–66.
- Pennell, M. W., L. J. Harmon, and J. C. Uyeda. 2013. Is there room for punctuated equilibrium in macroevolution? *Trends in Ecology and Evolution* 29:23–32.
- Pigliucci, M. 2009. An extended synthesis for evolutionary biology. *Annals of the New York Academy of Sciences* 1168:218–228.
- Pigliucci, M., and G. B. Müller. 2010. Evolution: the extended synthesis. MIT Press, Cambridge, MA.
- Provine, W. B. 1971. The origins of theoretical population genetics. University of Chicago Press, Chicago.
- . 1985. Adaptation and mechanisms of evolution after Darwin: a study in persistent controversies. Pages 825–866 *in* D. Kohn, ed. *The Darwinian heritage*. Princeton University Press, Princeton, NJ.
- . 1986. Sewall Wright and evolutionary biology. University of Chicago Press, Chicago.
- Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. *Animal Production* 8:93–108.
- Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, New York.
- Schmalhausen, I. I. 1949. Factors of evolution: the theory of stabilizing selection. Blakiston, Toronto.

Simpson, G. G. 1944. *Tempo and mode in evolution*. Columbia University Press, New York.
 ———. 1984. *Tempo and mode in evolution*. Columbia University Press, New York.
 Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences of the USA* 108:6187–6192.
 Stebbins, G. L. 1950. *Variation and evolution in plants*. Columbia University Press, New York.
 Stinchcombe, J. R., A. F. Agrawal, P. A. Hohenlohe, S. J. Arnold, and M. W. Blows. 2008. Estimating non-linear selection gradients using

quadratic regression coefficients: double or nothing? *Evolution* 62: 2435–2440.
 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 of the USA* 108:15908–15913.
 Wade, M. J., and C. J. Goodnight. 1998. Perspective: the theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. *Evolution* 52:1537–1553.
 Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97–159.

Editor: Judith L. Bronstein



“It is not in one alone of these groups that we find associated together, in a natural morphological combination, giants and dwarfs, although only in one do we find the contrast in the present age of our globe. It is the family of *Physeteridae* (the sperm-whales) which furnishes us with the contrast in living forms; only giants are now living to represent the *Balænidæ* (the right-whales), and *Balænopteriidæ* (the fin-back whales), but in the miocene age, a species of a fin-back whale lived that when adult was not even as large as the *new born young* of the fin-backs now living.” From “The Sperm Whales, Giant and Pygmy” by Theodore Gill (*The American Naturalist*, 1871, 4:725–743).