ARTICLES

Phenotypic Evolution in the Fossil Record: Numerical Experiments

Bjarte Hannisdal

Department of the Geophysical Sciences, University of Chicago,
5734 South Ellis Avenue, Chicago, Illinois 60637, U.S.A.
(e-mail: bhannis@geosci.uchicago.edu)

ABSTRACT

Stratophenetic data document phenotypic changes in a fossil lineage and play a vital role in reconciling contemporary microevolution with long-term paleontological patterns. However, stratophenetic series represent multiscale geological and biological interactions, defying simple analysis and interpretation. A numerical model is presented that simulates stratophenetic series in shallow marine siliciclastic deposits. The model is driven by predictions of water depth, substrate properties, and sedimentation rate from a high-resolution sedimentary basin fill model. Species abundance is modeled as a probability density peaked with respect to environmental preferences. The Price equation is used to model phenotypic evolution based on phenotype-fitness covariance and drift. Preservation is a Poisson process controlled by population size, preservation probability, and sedimentation. Numerical experiments are used to investigate (1) the effects of sampling and depositional architecture on observed patterns and (2) the performance of various statistical tests in identifying evolutionary mode. As sample sizes decrease, the inaccuracy of sample mean values causes a stratophenetic pattern referred to as analytical stasis. Depositional architecture can cause nonrandom patterns through temporally irregular preservation, with the relative size and distribution of gaps being more important than the absolute size of gaps and overall completeness. For short series, statistical tests based on a random-walk null hypothesis lose power and should be abandoned in favor of a multidimensional approach. A model of the data is needed that can account for confounding factors, and all available information on time and environment as well as phenotypic data should be incorporated and analyzed jointly, with a greater emphasis on quantifying uncertainty.

Online enhancement: appendix.

Introduction

The ability of population-level microevolutionary processes to generate rapid evolutionary changes has been extensively documented (e.g., Grant and Grant 1995; Reznick et al. 1997; Hendry and Kinnison 1999). How these contemporary rates and patterns of microevolution play out over long evolutionary timescales, however, remains unknown, as exemplified by the often-cited discrepancy between the high rates observed in modern populations and the low rates observed in the fossil record (Jackson and Cheetham 1999; Jablonski 2000; Gingerich 2001; Kinnison and Hendry 2001). The fossil record is the only source of direct evidence for the manifestation of microevolution in deep time, but the strength of that evidence is a subject of debate.

Manuscript received May 27, 2005; accepted September 30, 2005.

Paleontological data on phenotypic evolution consist of morphological measurements from stratigraphic successions of fossil samples, referred to as stratophenetic series (Gingerich 1976*a*; Roopnarine et al. 1999). Although the use of stratophenetic series to address questions of evolutionary rate (tempo) and pattern (mode) has a long pedigree (Rowe 1899; Brinkmann 1929; Simpson 1944), it gained considerable momentum from punctuated-equilibrium theory (Eldredge and Gould 1972), which postulates that the rarity of observed gradualistic change in the fossil record is not an artifact of incompleteness but an expected result of the concentration of phenotypic change in geologically instantaneous speciation events. Subsequent research thus focused on the following questions of central importance to evolutionary biology and pa-

[The Journal of Geology, 2006, volume 114, p. 133–153] 0022-1376/2006/11402-0001$15.00

leobiology: can microevolutionary processes be extrapolated to reconcile rates and patterns in modern populations with large-scale evolutionary phenomena (Stanley 1979; Charlesworth et al. 1982; Schluter 2000; Arnold et al. 2001), and what are the predominant modes of phenotypic evolution (Erwin and Anstey 1995; Jackson and Cheetham 1999)? Our answers to these and related questions depend in part on our ability to reliably measure evolutionary rates and infer evolutionary patterns from paleontological data.

The occurrence, abundance, and morphological variation of fossils in a stratigraphic section are governed by the interaction of geological, ecological, evolutionary, and taphonomic processes. The relative importance and types of relations among these various factors will depend on the scale and resolution (temporal, spatial, and taxonomic) of the study, the geological setting, and the type of organism. Deficiencies in our current understanding of these interactions and of the relevant parameters thus pose a serious challenge for the analysis and interpretation of stratophenetic data.

This article describes a numerical model of within-lineage (nonbranching) phenotypic evolution on a 1-m.yr. scale for a marine, soft-sediment, benthic invertebrate, sampled bed-by-bed from deposits on a simulated siliciclastic shelf. The model is used to run numerical experiments investigating the effects of sampling error and depositional architecture on stratophenetic patterns. It is shown how small sample sizes bias observed stratophenetic patterns in favor of stasis, a simple yet underemphasized phenomenon with important implications for the documentation of modes of evolution. Examples are given of how the nature of sedimentary basin fill can create nonrandom patterns in stratophenetic series. Furthermore, the analytical problems associated with random-walk tests underscore the need for a more comprehensive multidimensional approach to quantifying evolutionary tempo and mode in the fossil record.

Model Outline

Holland (1995, 2000) used stratigraphic simulation model output coupled with a model of species occurrence to generate a synthetic fossil record, showing how sequence stratigraphy influences patterns of first and last occurrences of taxa, with implications for biostratigraphy and rates of taxonomic evolution. A similar strategy is adopted here, using the following model components to generate synthetic stratophenetic series (fig. 1): a sedimentary basin fill model predicting water depth, sediment

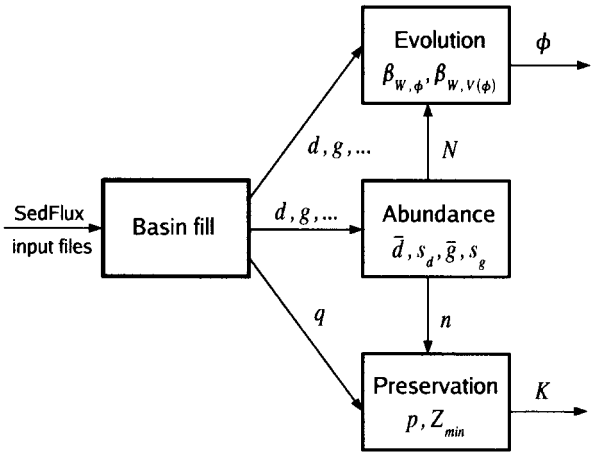

Figure 1. Outline of model components and their input/output relations. The basin fill model (SedFlux) takes a series of input files (defining basin dimensions, sediment properties, and process parameters) and outputs various seafloor properties, including water depth d and sediment grain size g, as well as depositional characteristics (e.g., sedimentation rate q). These output variables are used to drive models of (1) abundance, predicting population size N (the sum of population sizes per spatial bin n) based on a species' habitat preferences; (2) evolution, predicting population changes in phenotype ϕ in response to selection and drift; and (3) preservation, predicting the number of preserved fossils K.

substrate properties, and sedimentation rate along an onshore-offshore profile, according to user-defined sediment input, process parameters, and sea level change; a model of abundance predicting the distribution of individuals according to the species' habitat preferences and peak abundance; a model of phenotypic evolution predicting long-term changes in phenotypic population mean and variance in response to selection and drift; and a preservation model predicting the accumulation, time averaging, and burial (or erosion) of fossils as a function of population size, per capita preservation probability, and sedimentation rate. These model components are presented and discussed below.

Sedimentary Basin Fill. SedFlux is a comprehensive sedimentary process-response model designed to be coupled with river basin models to simulate the transport, deposition, reworking, and burial of sediment on continental margins (Syvitski and Hutton 2001). An important advance from earlier geometric stratigraphic simulation models (Paola 2000) is the ability to take into account natural variability in sediment flux (e.g., seasonality) as well as rare, catastrophic events (e.g., slope failures)

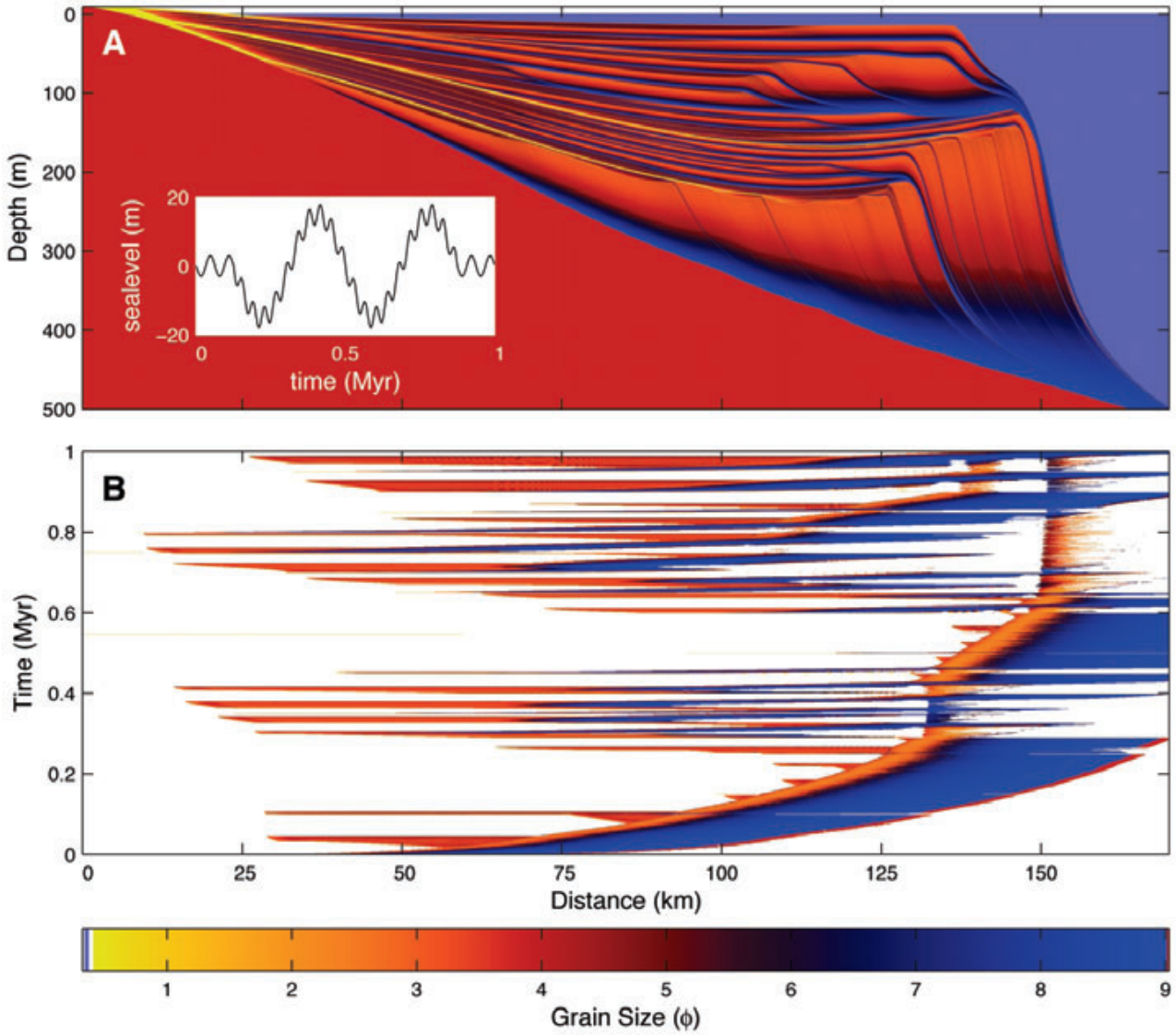

Figure 2. Simulated basin fill from the model SedFlux. Output is bin-averaged grain size, ranging from coarse sand (*yellow*) to clay (*blue*). Spatial bins are 50 m horizontal and 0.2 m vertical. *A*, Onshore-offshore cross section showing sediment distribution and architecture of the marine deposits. Note vertical exaggeration. *Inset*, hypothetical sinusoidal sea level curve with 40-m amplitude and minor fluctuations but no net trend. Because of isostasy, the deposits are slightly curved, and the original shelf topography is flattened. *B*, Chronostratigraphic (Wheeler) diagram showing grain size distribution on the seafloor at each time step, with white areas representing periods of nondeposition or erosion. Deactivation of slope failures and debris flows has reduced slope erosion and deep-water deposition. The output has been temporally rescaled.

and to capture the dynamic interactions of various sedimentary processes as boundary conditions change (Morehead et al. 2001; O'Grady and Syvitski 2001; Syvitski and Bahr 2001; Hutton and Syvitski 2004). SedFlux combines several individual process components, including models for bedload, suspended load (plumes), turbidity currents, storms, slope failure and debris flows, and compaction. This article uses the two-dimensional version of SedFlux. Input includes user-defined basin dimensions, sea level change, tectonic activity, sedimentary grain properties, and various process parameters (e.g., river input, storms, turbidity currents) as well as temporal (daily to decadal), horizontal (meter-scale), and vertical (centimeter-scale) resolution. Output includes bin-averaged grain size, wa-

ter depth, depositional thickness, porosity, bulk density, sediment age, and other sediment properties.

Figure 2 shows the output of a SedFlux run, using a hypothetical sea level curve to generate a basin history with depositional heterogeneity but no overall sea level or facies trend. Each spatial bin covers 50 m horizontally and 0.2 m vertically. Figure 2*A* shows a cross section of the basin with the grain size data plotted as a function of depositional thickness, including the underlying basement and the current sea level. Figure 2*B* shows the same data plotted as a function of time (i.e., a chronostratigraphic/Wheeler diagram), where white areas represent periods of nondeposition or erosion. It shows clear onlap-offlap patterns with rapid lateral movement of the shoreline (because of the shallow gradient) during transgressive-regressive cycles. The two main cycles show some differences in sediment distribution and patterns of sediment deposition and bypass, caused by differences in existing seafloor topography.

Because of the trade-off between spatial and temporal resolution, SedFlux was run with 10-yr time steps over 10^4 yr but with a sea level curve and sedimentation rate that resulted in unrealistically rapid lateral facies shifts and deposition for such short time steps. This was done to facilitate a scaling of the SedFlux output such that the temporal scope would increase to 10^6 yr, which is more relevant to paleontological tempo and mode studies. Each time step is thus assumed to represent 10^3 yr, which then equals the minimum scale of time averaging in each fossil sample and results in overall sedimentation rates of 0.5–0.1 m/k.yr. This rescaling thus renders the SedFlux output consistent with empirical measures of within-bed paleontological resolution (hundreds to thousands of years; see Kowalewski and Bambach 2003) and rates of deposition on terrigenous shelves measured over 10^6 yr (Sadler 1981). To further reduce computational cost, the most intensive SedFlux processes (slope failures, slumps, and debris flows) were turned off, which reduces the amount of slope erosion and deep-water deposition. The input parameter values are included in "SedFlux Input Parameter Values" in the appendix available in the online edition or from the *Journal of Geology* office. An important future goal is to run SedFlux over 10^6 yr to avoid scaling the output by using the full suite of processes and more realistic sea level curves in order to simulate a particular margin.

For the purpose of simulating stratophenetic series, SedFlux is considered a "black box" in the sense that it is run independently to generate sedimentary basin fill predictions, which are subsequently used to drive the models of abundance, evolution, and preservation. The basin fill model thus provides physical controls on the spatiotemporal heterogeneity characteristic of ecological/evolutionary systems (Plotnick 1996) but without dynamical coupling (i.e., there is no organism-sediment feedback). This approach is analogous to the way some earth scientists use general circulation model predictions to drive models of fluvial response and geomorphology (e.g., Boogart et al. 2003) or to explore vegetation responses to climate change (e.g., Shafer et al. 2001).

Habitat Preference and Abundance. Marine soft-bottom invertebrates are characterized by highly variable abundance and patchy distribution in space and time (Gray 1981; Parsons et al. 1984; Lenihan and Micheli 2001). Although many physical, chemical, and biological factors influence the distribution of marine benthic organisms, the model presented here assumes that habitat preference and abundance are associated with water depth and substrate sediment characteristics (e.g., average grain size, bulk density, percentage clay) either directly, for physical/biomechanical reasons, or indirectly, as a result of disturbance or nutrient availability (Levinton 1982; De Rijk et al. 2000; Sisson et al. 2002; Van Hoey et al. 2004). It is further assumed that the association between abundance and environment is Gaussian, with a single peak representing an optimal combination of environmental properties (Whittaker 1970; Brown 1995), although this assumption can be relaxed without major effects on the results presented below. This is a multivariate version of the approach taken by Holland (1995, 2000), who modeled the probability of occurrence as a Gaussian density using three parameters: preferred depth, depth tolerance, and peak abundance.

Let d denote water depth and g denote grain size. Let $\overline{d}$ be the species' preferred depth, s_{d} the depth tolerance, $\overline{g}$ the preferred grain size, and s_{g} the grain size tolerance. Note that, in general, indexes (temporal, spatial, and individual) are dropped from the equations. If we think of $\overline{d}$, $\overline{g}$ as analogous to the first moments (means) and s_{d}^2, s_{g}^2 as analogous to the second moments (variances) of probability densities, they describe Gaussian response curves along two environmental gradients. If we assume that the species' response to each environmental variable is uncorrelated, a bivariate Gaussian density function

(Wackerly et al. 2002) for the probability of occurrence is given by

$$f(d, g) = \frac{1}{2\pi s_d s_g} \exp\left(-\frac{1}{2}\left[\frac{(d-\overline{d})^2}{s_d^2} + \frac{(g-\overline{g})^2}{s_g^2}\right]\right). \quad (1)$$

The parameters $\overline{d}$, $\overline{g}$, s_d^2, and s_g^2 are thus used to control the environmental sensitivity and abundance distribution of a simulated benthic species. Figure 3*A* shows an example of a density defined by equation (1) for a particular set of parameter values. The number of individuals n in each spatial bin x is found by scaling the peak of the density function (fig. 3*A*) to the maximal per-bin abundance n_{max}, and the total population size is then the sum over all bins:

$$N = \sum^{x}\left[n_{max}\frac{f(d_x, g_x)}{f(\overline{d}, \overline{g})}\right]. \quad (2)$$

When applied to the basin fill history from SedFlux, the above model allows a simulated benthic species to track its preferred habitat in space and time. Although the underlying probability density is a symmetric Gaussian, the realized abundance distributions along the seafloor can look drastically different at different times as a result of variability in substrate properties and in the bathymetric profile in response to sea level changes (fig. 3*B*–3*F*). This matches the real-

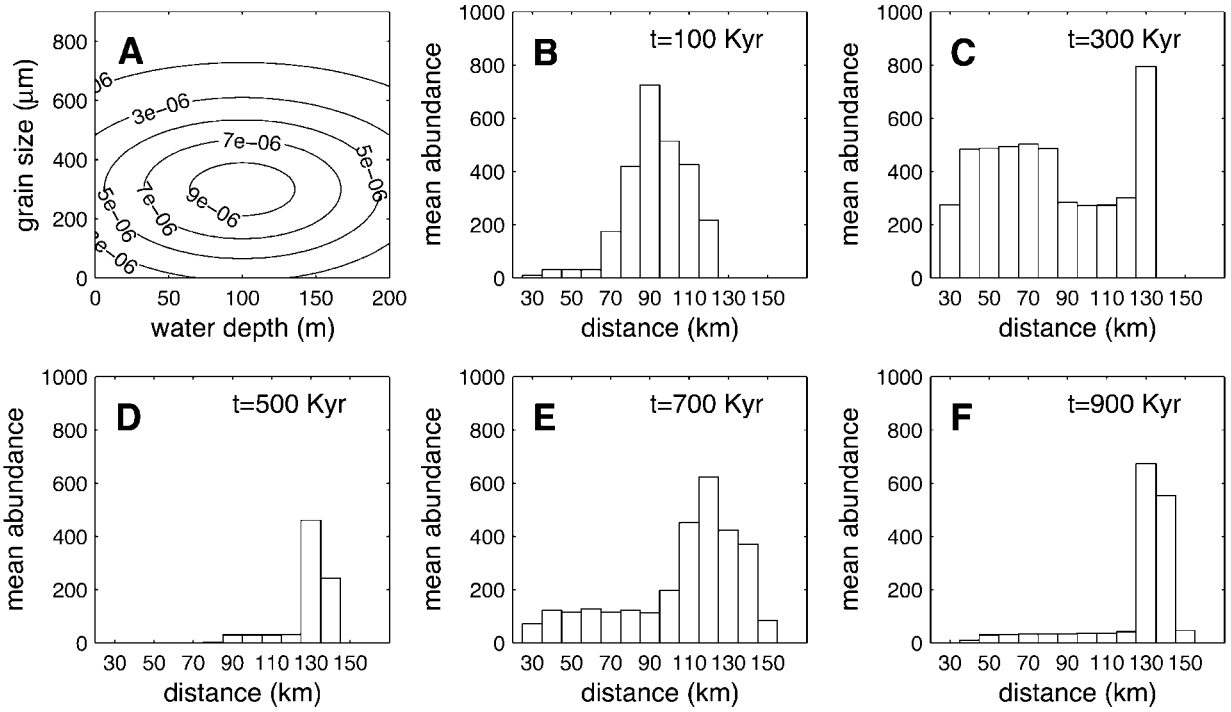

Figure 3. Model of habitat preference and abundance. *A*, Gaussian density representing the probability of occurrence of a benthic species, here controlled by two environmental variables: water depth ($\overline{d}$ = 100 m, s_d = 80 m) and grain size ($\overline{g}$ = 300 μm, s_g = 200 μm). The peak of the density function is scaled to the maximum per-bin abundance. *B*–*F*, Realized abundance distributions across the basin profile at various times throughout a model run, using the basin fill history from figure 2 and the Gaussian model in *A*. Each bar represents the mean abundance of 200 spatial bins (10 km) along the seafloor. Because of variations in the bathymetric gradient and sediment grain size distribution in response to sea level changes, the abundance distributions look different at different times, ranging from relatively symmetric (*B*) to highly skewed (*E*), with a mode that is either variable in both magnitude and position or polymodal (*C*). Even for a species with broad depth tolerance, the actual area occupied at any given time can be a limited portion of the basin (*D*).

world observation that a given benthic marine species does not have the same spatial distribution pattern everywhere, which is why the use of benthic fossils as paleobathymetric indicators must be based on assemblages rather than single species (e.g., Murray 1991).

Phenotypic Evolution. For any variable influenced by a set of predictors, the total change in the variable can be partitioned into two components (Frank 1997): (1) change in the frequency of predictors and (2) difference in the effect of predictors in the context of the changed population. In the case of natural selection, these two components are represented by fitness (differential reproductive success as a function of trait value) and transmission (the degree to which offspring resemble parents), respectively. This relationship can be used to formulate the evolutionary change in a phenotypic trait as the so-called Price equation (Price 1970; Frank 1997; Rice 2004):

$$\Delta\overline{\phi} = \frac{1}{\overline{W}}[\mathrm{Cov}(W, \phi) + \mathrm{E}(W\overline{\delta})], \tag{3}$$

where ϕ is some measurable phenotypic trait with population mean $\overline{\phi}$, W is the number of descendants per individual (fitness) with population mean $\overline{W}$, $\overline{\delta}$ is the difference in the mean value of ϕ between ancestors (parents) and descendants (offspring), and Cov and E denote covariance and expectation, respectively (see "The Price Equation" in the appendix).

Equation (3) is not restricted to time steps of one generation and can be applied to any phenotypic value (or function thereof) that can be defined in both ancestor and descendant. The first term on the right-hand side of equation (3), $1/\overline{W}\,\mathrm{Cov}(W, \phi)$, is the selection differential, which represents the change as a result of differential survival and reproduction. If we rewrite this term as $1/\overline{W}\beta_{W,\phi}\,\mathrm{Var}(\phi)$, we can interpret phenotypic evolution by selection as being determined by the linear regression of fitness on phenotype, $\beta_{W,\phi}$, regardless of the shape of the joint distribution of fitness and phenotype.

The second term, $1/\overline{W}\mathrm{E}(W\overline{\delta})$, is generally used to account for change resulting from processes involved in reproduction (e.g., recombination), which is important when modeling evolution over a generation (Rice 2004). It is assumed that on the timescale considered here, such transmission effects can be ignored, and we instead use the second term to account for the direct effects of environment on phenotype (phenotypic plasticity or ecophenotypy). However, ecophenotypy is not implemented in the numerical experiments described in this article (see "The Price Equation" in the appendix for more details).

Another important component of phenotypic evolution is the change in phenotypic variance. We can use the fact that $\mathrm{Var}(\phi)$ is equal to the mean of $(\phi - \overline{\phi})^2$ to simply substitute $(\phi - \overline{\phi})^2$ for ϕ in equation (3):

$$\Delta\,\mathrm{Var}(\phi) = \frac{1}{\overline{W}}\beta_{W,(\phi-\overline{\phi})^2}\,\mathrm{Var}((\phi - \overline{\phi})^2). \tag{4}$$

The two linear regression parameters $\beta_{W,\phi}$ and $\beta_{W,(\phi-\overline{\phi})^2}$ can thus be used to model directional selection ($\beta_{W,\phi} \neq 0$, leading to a change in mean phenotype), stabilizing selection ($\beta_{W,(\phi-\overline{\phi})^2} < 0$, leading to a reduction phenotypic variance), and disruptive selection ($\beta_{W,(\phi-\overline{\phi})^2} > 0$, leading to an increase in phenotypic variance).

In a finite population, a nonzero covariance between fitness and phenotype can also result from sampling descendants from the ancestral distribution of fitness values associated with each value of ϕ, which manifests itself as drift. If the population does not grow or decline as a result of differential reproductive success, a valid fitness distribution for selectively neutral variation is a Poisson with mean $\overline{W} = 1$ (and thus a variance of 1), which simplifies the variance in phenotypic change to the following expression (Rice 2004):

$$\mathrm{Var}(\Delta(\phi)) = \frac{\mathrm{Var}(\phi)}{N}. \tag{5}$$

Drift is thus inversely proportional to population size N under these conditions. Figure 4 shows an implementation of this model and the effects of varying the three main parameters, $\beta_{W,\phi}$, $\beta_{W,(\phi-\overline{\phi})^2}$, and N. Flexibility is added by allowing parameters to vary through time, for example, to simulate fluctuating selection pressures. For the numerical experiments described in this article, however, $\beta_{W,\phi}$ and $\beta_{W,(\phi-\overline{\phi})^2}$ (but not N) are held constant throughout model runs. A range of different evolutionary modes can be generated using these parameters, including stasis (by stabilizing selection or fluctuating selection), directional trends, and punctuations (which may be caused by drift during population bottlenecks or pulsed selection), and several processes can operate simultaneously (e.g., drift superimposed on both directional and stabilizing selection).

Preservation. The number of individuals per

model bin that are preserved as fossils, K, is determined by (*a*) per-bin population size n, where each individual represents a binomial (success-failure) trial; (*b*) per-individual preservation probability p, representing the intrinsic fossilization potential of the organism as well as the probability of collection; and (*c*) per-bin sedimentation rate q (the difference in depositional thickness between time steps), which exerts a basic control on preservation through erosion versus deposition but can also be used to modulate preservation probability by the amount of deposition (increased chance of burial). Given that n is generally very large and p is very small, we can model this as a Poisson process with density function

$$f(K, \lambda) = \frac{e^{-\lambda}\lambda^{K}}{K!}. \tag{6}$$

The rate parameter λ (the expected number of instances of preservation) is expressed as

$$\lambda \equiv \begin{cases} 0 & \text{for } q < Z_{\min}, \\ np[(Z_{\min} - q)/Z_{\min}] & \text{for } q \geq Z_{\min}, \end{cases} \tag{7}$$

where $Z_{\min}$ is a negative thickness value representing the amount of erosion necessary to remove the entire population (and below which previously preserved fossils may be eroded). The implementation of the preservation model tracks both the age and the stratigraphic position of each preserved fossil, giving a measure of the magnitude of time averaging in the fossil samples (beyond the limiting resolution of the 10^3-yr time steps) and its stratigraphic distribution. However, the analytical effects of time averaging, such as inflation of the phenotypic variance of samples, are negligible in the model runs described below. This is consistent with the results of empirical and analytical studies of the effect of time averaging on morphological data in the fossil record (Bell et al. 1985; Bush et al. 2002; Hunt 2004*a*, 2004*b*).

Synthetic stratophenetic series are thus generated by combining a simulated basin history from SedFlux with the models outlined above. In addition to environmental and depositional variables, output includes population size, phenotypic mean and variance, preserved sample size, and sample mean and variance. Sampled phenotypic values are referred to as stratophenetic series, whereas the true population values are referred to as the microevolutionary series.

Numerical Experiments

Analyses and interpretations of stratophenetic series are faced with powerful confounding factors that can be crudely categorized into (1) sampling error, (2) depositional architecture, and (3) ecophenotypic variation. All three presumably influence observed stratophenetic series in some way and thus our ability to correctly infer microevolutionary patterns. However, lack of necessary information (such as estimates of geographic variation and sequence stratigraphy) makes it difficult to assess their importance from published empirical case studies. The numerical forward model presented above can predict the behavior of paleontological data on spatiotemporal scales that are difficult to visualize and understand. This section describes the results of numerical experiments used to investigate some of the factors influencing stratophenetic patterns, focusing on the qualitative and quantitative effects of sampling and depositional architecture.

Depositional history is the same for all experiments (fig. 2), and comparisons are made among different locations ("measured sections") across the basin as well as between the stratophenetic and the microevolutionary series. All model runs assume that a time step of 10^3 yr is sufficiently greater than the generation time to cause gene flow to erase spatial population structure within the basin, effectively modeling a single, panmictic population. A related assumption is that the species is endemic to the basin, which means that the observed phenotypic variation is not affected by immigration from other regions. These are strong assumptions, and the effects of their violation will be an important topic for future work. Although stratophenetic series can be recovered from widely spaced localities, the actual area occupied by the population at any given time may be only a limited portion of the basin, even for a species with broad environmental tolerance (fig. 3). The species tracks its preferred environment, which shifts up and down the onshore-offshore profile, leaving a diachronous fossil record.

In order to compare the results for stratophenetic and microevolutionary series among different locations and among different experiments consistently, three statistical tests have been applied: the runs test, the scaled maximum test, and the Hurst exponent test. All of these have been proposed to quantify evolutionary tempo and mode in fossil data, based on testing for significant deviation from a random-walk null hypothesis (Raup and Crick 1981; Bookstein 1987; Roopnarine 2001). The runs

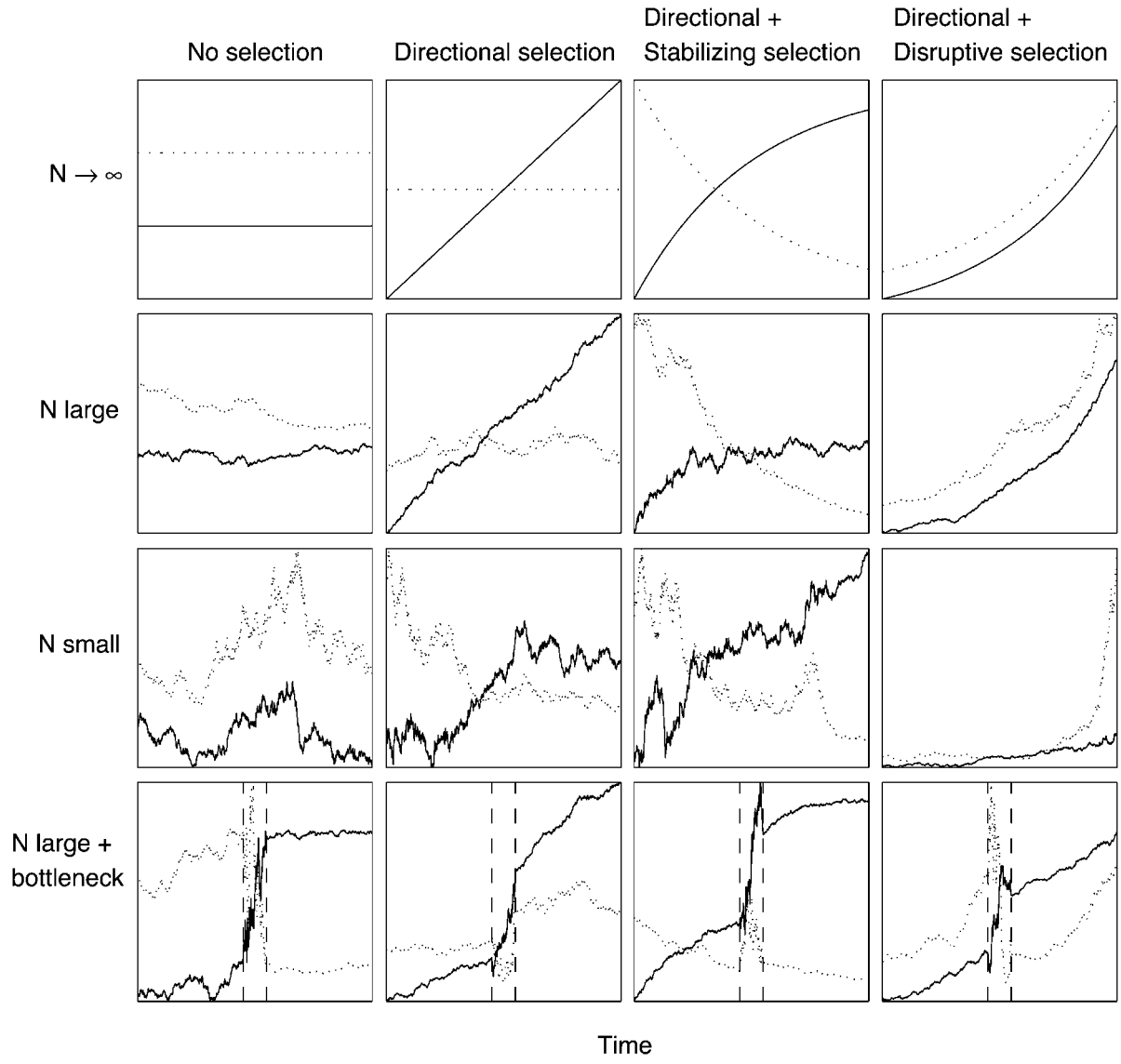

Figure 4. Model of phenotypic evolution based on the Price equation. Each panel shows the change over time (1000 steps) in the mean phenotypic value (*solid line*, starting at 0) and the phenotypic variance (*dotted line*, starting at 1) of the population. The top row of panels depicts different scenarios based on the values of the two regression parameters, from left to right: no selection ($\beta_{W,\phi} = 0$, $\beta_{W,(\phi-\bar{\phi})^2} = 0$), directional selection ($\beta_{W,\phi} > 0$, $\beta_{W,(\phi-\bar{\phi})^2} = 0$), directional selection with stabilizing selection ($\beta_{W,\phi} > 0$, $\beta_{W,(\phi-\bar{\phi})^2} < 0$), and directional selection with disruptive selection ($\beta_{W,\phi} > 0$, $\beta_{W,(\phi-\bar{\phi})^2} > 0$). The two middle rows are for the same scenarios but with decreasing population size, resulting in stronger drift. The bottom row shows the effect of introducing a population bottleneck (reducing the population 10-fold over a period of time indicated by dashed lines), which can generate a rapid shift in the mean value and a spike in and/or depletion of variance.

test operates on the sequence of positive and negative increments in a series, comparing the number of runs (consecutive steps with equal sign) in a series with that expected from a random walk. The scaled maximum test and the Hurst exponent compare the maximum change or rate of change to that expected from a random walk (see "The Runs Test," "The Scaled Maximum Test," and "The Hurst Exponent" in the appendix for details). The tests are applied to the stratophenetic series from each location as well as to the microevolutionary series. The latter is subsampled to match the length

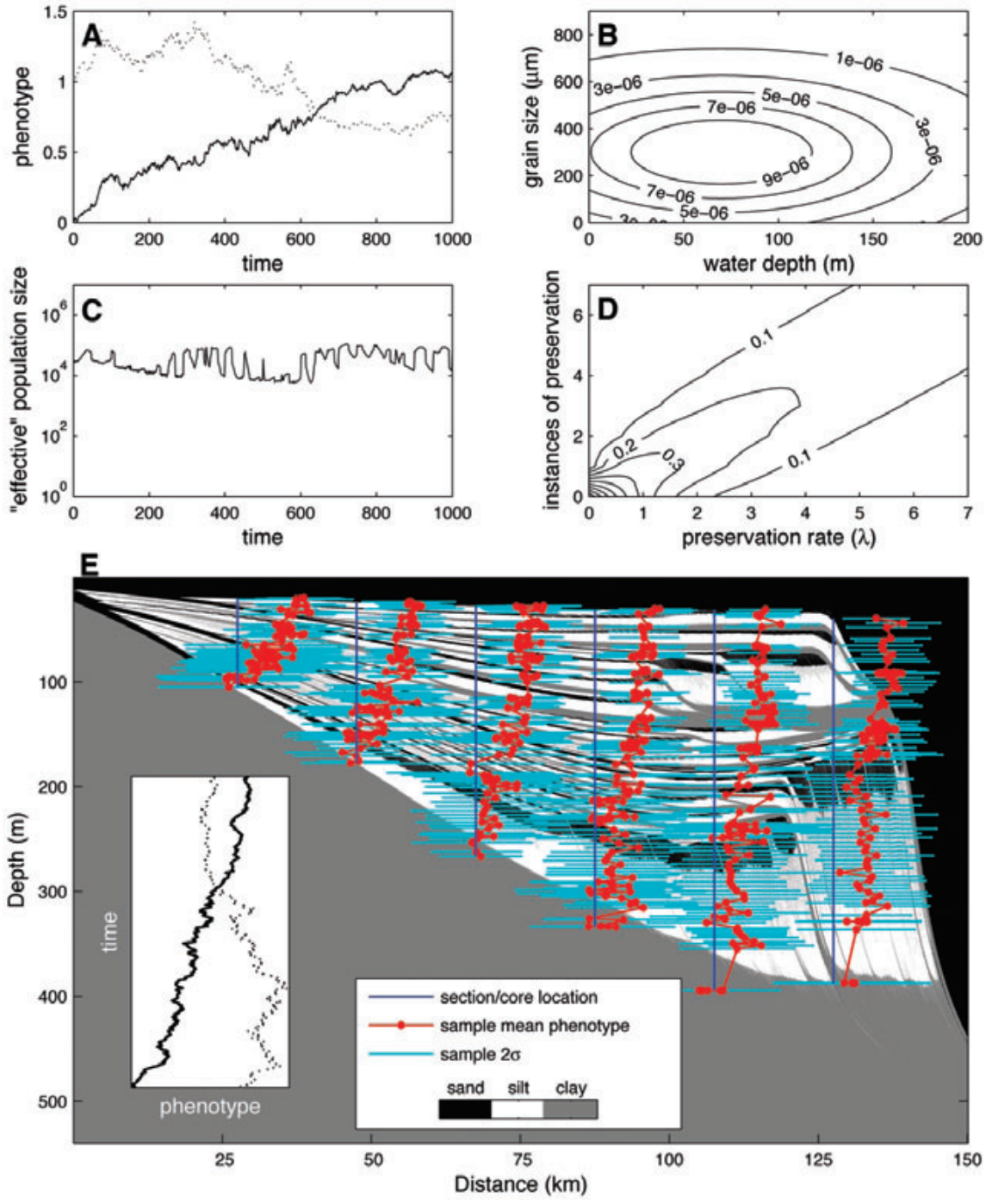

Figure 5. Numerical experiment with directional selection and small sample sizes. *A*, Phenotypic evolution with directional selection, based on parameters $\beta_{W,\phi} = 0.001$, $\beta_{W,(\phi-\bar{\phi})^2} = 0$, and N from *C*. Solid line is for phenotypic mean; dotted line is for variance. *B*, Gaussian density based on preferred depth = 70 m, depth tolerance = 70 m, preferred grain size = 300 μm, and grain size tolerance = 200 μm. *C*, Population size (in this case equal to effective population size) through time, based on the model in *B*. *D*, Poisson density function used to determine preservation based on λ as a function of the abundance in each model bin. *E*, Stratophenetic series plotted against synthetic basin fill from figure 2. Sample means (*red*) and error bars (*cyan*) are scaled to make the stratophenetic patterns easily visible and have to be projected back onto the blue line ("core") for the samples to be in original stratigraphic position. *Inset*, microevolutionary series from *A*. Small sample sizes cause the sample means to lose accuracy, resulting in spurious patterns of stasis (see table 2).

Table 1. Test Results for Drift Experiment with Small Sample Sizes

	Core					
	1	2	3	4	5	6
Position (km)	27.5	47.5	67.5	87.5	107.5	127.5
Series length	23	42	72	80	121	121
Sample sizes	20–30	20–21	20–44	20–55	20–49	20–58
Runs test:						
Z	3.06	2.06	3.23	3.30	5.17	3.40
Mode	Stasis	Stasis	Stasis	Stasis	Stasis	Stasis
Z_{true}	−.874	−.124	−1.07	1.26	−.685	−.135
Mode	Random	Random	Random	Random	Random	Random
Scaled maximum test:						
x	.435	.369	.396	.435	.287	.405
Mode	Stasis	Stasis	Stasis	Stasis	Stasis	Stasis
x_{true}	.892	.928	.897	.863	.948	1.36
Mode	Random	Random	Random	Random	Random	Random
Hurst exponent:						
H	.168	.105	.0685	.106	.101	.192
Null mean	.511	.506	.495	.493	.495	.503
2.5%	−.0716	.111	.117	.134	.226	.218
97.5%	1.06	.920	.833	.859	.793	.786
Mode	Random	Stasis	Stasis	Stasis	Stasis	Stasis
H_{true}	.700	.345	.450	.358	.370	.463
Mode	Random	Random	Random	Random	Random	Random

Note. For positions of cores, see figure 5. Subscript "true" refers to test statistic calculated on the microevolutionary series. Cutoff values for runs test are ± 1.96; cutoff values for scaled maximum test are 0.62 and 2.25. Null mean and percentiles calculated from a distribution of H estimates for 500 random-walk realizations of the same length as the observed series. For details, see "The Runs Test," "The Scaled Maximum Test," and "The Hurst Exponent" in the appendix available in the online edition or from the *Journal of Geology* office.

of each stratophenetic series by sampling at points in time that (*a*) coincide with sample ages, (*b*) are chosen randomly, or (*c*) are evenly spaced. Results of the tests are labelled "random" (failure to reject a random-walk null hypothesis), "directional" (observed data show more net change than expected for a random walk), or "stasis" (less net change than expected for a random walk). In addition, a fourth test, described by Charlesworth (1984) and Cheetham (1986), is applied to instances of punctuated stratophenetic series. It compares estimated mean and variance in the rate of change within an ancestral species with that measured between ancestor and descendant species (i.e., pre- and postpunctuation series), with gradualism as a null hypothesis (see "The Charlesworth-Cheetham Test" in the appendix).

Effects of Sampling Error. The following three aspects of sampling affect the information content of stratophenetic series: (1) sample size, which is the number of individuals per sample; (2) series length, which is the number of samples constituting the series; and (3) sampling at irregular time intervals. Series length varies among locations within the basin primarily because of depositional architecture and has serious implications for the analysis of stratophenetic series because of its effect on the statistical power of tests, as the following experiments show. All model runs have preservation probabilities set to produce series lengths in the range of 20–150, similar to (and in most cases greater than) those obtained in empirical studies, which range from <10 (e.g., McCormick and Fortey 2002) to around 100 (e.g., Malmgren et al. 1983). Temporal irregularity of sampling is influenced primarily by depositional architecture, and its effects are closely linked to uncertainty in the age model. This section focuses on the effects of sample size in terms of the accuracy with which the population mean is estimated. Sample sizes encountered in empirical stratophenetic series are highly variable and may range from 2 to 420 specimens in a single study (Girard et al. 2004), although sample sizes are rarely >75 and commonly <25 specimens.

In order to isolate the effects of sample size as defined above, it is necessary to switch off other confounding factors. To avoid the influences of inaccurate sample age estimates, the age model is set to true time (i.e., sample ages are known with certainty), and to control for temporal irregularity of sampling, the microevolutionary series is subsampled at exactly the same points in time as the fossil samples. In addition, there is no ecophenotypic variation. The minimum sample size is 20 specimens.

Figure 5 shows a model run with directional selection and small sample sizes (20–58 specimens

Table 2. Test Results for Directional Selection Experiment (Fig. 5)

	Core					
	1	2	3	4	5	6
Position (km)	27.5	47.5	67.5	87.5	107.5	127.5
Series length	65	80	102	124	144	135
Sample sizes	20–38	20–46	20–51	20–46	20–48	20–59
Runs test:						
Z	2.11	2.39	3.71	4.34	4.11	4.87
Mode	Stasis	Stasis	Stasis	Stasis	Stasis	Stasis
Z_{true}	−.269	−.0179	.127	−.783	−.212	2.16
Mode	Random	Random	Random	Random	Random	Directional
Scaled maximum test:						
x	.437	.506	.500	.388	.447	.509
Mode	Stasis	Stasis	Stasis	Stasis	Stasis	Stasis
x_{true}	2.58	2.02	3.39	3.22	2.58	4.16
Mode	Directional	Random	Directional	Directional	Directional	Directional
Hurst exponent:						
H	.189	.115	.292	.220	.273	.166
Null mean	.490	.509	.506	.503	.510	.496
2.5%	.128	.145	.191	.232	.217	.187
97.5%	.853	.846	.809	.786	.777	.785
Mode	Random	Stasis	Random	Stasis	Random	Stasis
H_{true}	.946	.898	.873	.833	.833	.787
Mode	Directional	Directional	Directional	Directional	Directional	Directional

Note. For positions of cores, see figure 5. Subscript "true" refers to test statistic calculated on the microevolutionary series. Cutoff values for runs test are ± 1.96; cutoff values for scaled maximum test are 0.62 and 2.25. Null mean and percentiles calculated from a distribution of H estimates for 500 random-walk realizations of the same length as the observed series. For details, see "The Runs Test," "The Scaled Maximum Test," and "The Hurst Exponent" in the appendix (available in the online edition or from the *Journal of Geology* office).

per sample). Phenotypic evolution (fig. 5*A*) is a directional trend in the mean and moderate drift, with a slight decrease in variance. The species prefers 70-m water depth but has a broad depth tolerance (fig. 5*B*) and a more restricted grain size preference, which causes population size to fluctuate but remain relatively large through basin history (fig. 5*C*, note log scale). Very large population sizes tend to reduce the effect of drift (i.e., smoothing out the phenotypic mean trajectory), and total population size is therefore scaled down ($<10^5$) to emulate a smaller "effective" population size (i.e., the number of individuals in a population who contribute offspring to the next generation). Abundance in each model bin along the seafloor is the time-varying component of preservation rate λ, which is the controlling parameter of the preservation probability density (fig. 5*D*). Stratophenetic series (fig. 5*E*) are composed of fossil sample values plotted in stratigraphic position.

Although drift is always present in the phenotypic evolution model, an evolutionary mode of purely random drift is here generated by setting a constant, small "effective" population size of 10^3. This is unrealistic but helps avoid unwanted effects of fluctuating population size, the goal being to make the microevolutionary series equivalent to a random walk for the purpose of hypothesis testing. The output from the pure-drift experiment is not plotted, but table 1 shows the results of applying the random-walk tests to the output. Although the microevolutionary series is correctly classified as random, all three tests classify the stratophenetic patterns as stasis. Significance is remarkable, given the low statistical power characterizing the tests. For instance, the most proximal series (core 1, at 27.5 km) is very short (series length = 23), undermining the assumptions of both the runs test and the scaled maximum test, and the Hurst exponent confidence interval for a random walk of the same length is correspondingly very large (in this case, the failure to reject the null is not surprising because the 0.95 confidence interval exceeds the expected theoretical limits of the Hurst exponent [0, 1]!).

Table 2 shows the test results for the directional selection experiment in figure 5, which are very similar to the results for drift: the stratophenetic patterns are classified as stasis despite relatively long series and the appearance of a directional trend. If the strength of directional selection is increased, all tests correctly identify the microevolutionary series (which approaches a straight line) as directional, but the stratophenetic series are still misclassified, although more frequently as random rather than stasis (results not presented). Further-

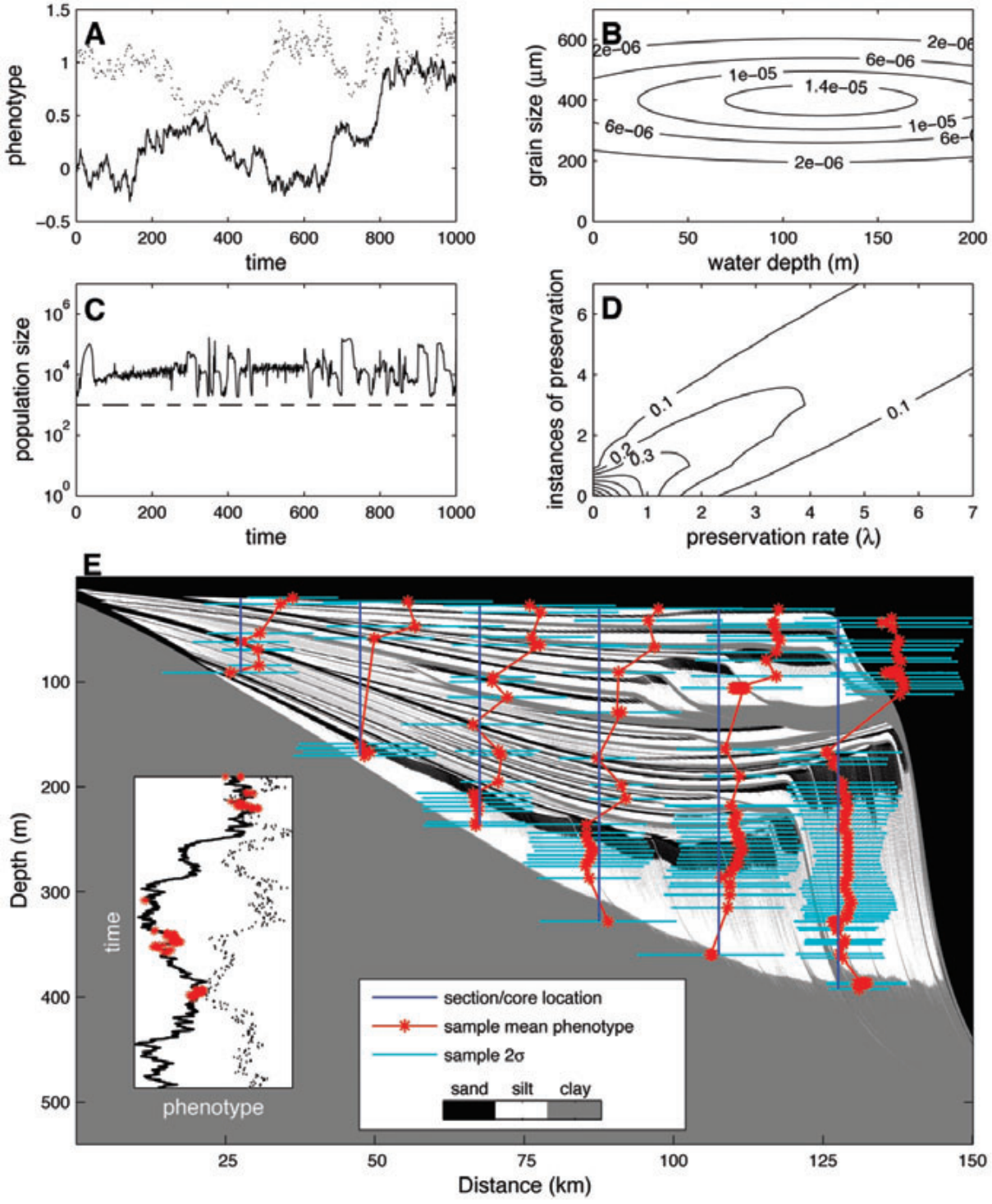

Figure 6. Numerical experiment showing the effect of depositional architecture. *A*, Phenotypic evolution as random drift, based on parameters $\beta_{W,\phi} = 0$, $\beta_{W,(\phi-\bar{\phi})^2} = 0$, and N from *C*. Solid line is for phenotypic mean; dotted line is for variance. *B*, Gaussian density based on preferred depth = 120 m, depth tolerance = 100 m, preferred grain size = 400 μm, and grain size tolerance = 100 μm. *C*, Population size through time based on the model in *B*. Solid line is for total population size; stippled line is for "effective" population size, which in this case is held constant and small. *D*, Poisson density function used to determine preservation based on λ as a function of the abundance in each model bin. *E*, Stratophenetic series plotted against synthetic basin fill from figure 2. Sample means (*red*) and error bars (*cyan*) are scaled to make the stratophenetic patterns easily visible and have to be projected back onto the blue

more, the results in tables 1 and 2 can be considered best-case scenarios in the sense that there is no ecophenotypy and the sample ages are known with certainty. If a linear age model, stratigraphic position, or sample number is used instead to arrange the samples with respect to time, the error rate increases.

Familiarity with the statistical behavior of small and "noisy" data sets will give most paleontologists an intuitive appreciation for why small sample sizes give the impression of stasis: as the accuracy of sample means decreases, mean values will fluctuate within the population variance (alternating between over- and underestimation), and the variability within the series will overwhelm temporal patterns or net change. This tendency toward Type I error (erroneous rejection of the null) in favor of stasis as a result of sampling error is referred to here as analytical stasis. Although conceptually trivial, analytical stasis may have serious implications for the documentation of tempo and mode in the fossil record.

Effects of Depositional Architecture. An important geological control on the nature of stratophenetic data is depositional architecture, that is, the configuration of unconformity-bounded sedimentary bodies constituting the basin fill. Depositional architecture in reality represents a combination of several biases, including (*a*) temporally irregular sampling (the main topic here) related to upsection depositional unsteadiness, environmental changes, and age model uncertainty but also (*b*) geographic and ecophenotypic variation and (*c*) the bottleneck problem. Experiments described in this section have sample means that reliably estimate the population mean values. In the model, sample size effects are minimized by boosting preservation probability and analyzing only very large samples (i.e., $K > 500$). To control for effects of series length, the microevolutionary series is subsampled randomly to match the length of the stratophenetic series.

Figure 6 shows a model run with phenotypic evolution as purely random drift, achieved by setting a constant small effective population size (fig. 6*C*). Qualitatively, the stratophenetic patterns vary among different locations within the basin (fig. 6*E*), from highly incomplete series with some net directionality (proximally), to series that look more oscillating (medially), to the distal section showing, basically, two static series separated by a jump. Along the onshore-offshore profile, stratigraphic sections capture different segments of history (e.g., Perlmutter and Plotnick 2003), and thus stratophenetic series capture different segments of the microevolutionary time series.

Quantitatively, the effect of temporally irregular sampling on stratophenetic series depends on the age model. Although the runs test and the scaled maximum test are insensitive to how time is measured, the Hurst exponent test is an explicit function of time and is sensitive to the age model. For example, if all sample ages are known with certainty, then most of the stratophenetic series in figure 6 would be classified as random, although the runs test classifies both distal series as stasis (results not presented). In contrast, if a linear age model is used (i.e., endpoint sample ages are known, and remaining samples are interpolated based on stratigraphic position), then the Hurst exponent test classifies core 5 (at 107.5 km) as stasis; the same result is obtained using stratigraphic position uncalibrated to absolute time. This suggests that elements of sampling other than sample size can generate analytical stasis.

The most distal stratophenetic series is separated into two subseries by a very pronounced jump, or punctuation, where each subseries represents dense sampling of populations closely spaced in time during relatively short periods of high sedimentation rate, thus capturing small segments of the microevolutionary series widely separated in time. This is illustrated in figure 6*E* (*inset*), where the sample means of the distal series are superimposed on the microevolutionary series. Note that the lack of slope failure processes will have oversimplified the slope stratigraphy, thus inflating the completeness of the distal sections to some degree. The significance of the punctuation was assessed using the Charlesworth-Cheetham test (table 3). Assuming a perfect age model (time), the series differs significantly in variance within and among subseries but does not differ significantly in the mean rate of change. The same result is obtained with a linear age model. However, for an age model based on stratigraphic position (thickness), both tests result in rejection of a gradualistic null hypothesis at the

line ("core") for the samples to be in original stratigraphic position. *Inset*, the sample means of the most distal series superimposed on the microevolutionary series. Sample means are accurate estimates of the population mean, and depositional architecture causes the observed patterns to vary among locations (see table 3).

Table 3. Results of the Charlesworth-Cheetham Test on the Most Distal Series in Figure 6

Age model	$\Delta X_W/T_W$	s^2_W	$\Delta X_A/\Delta T_A$	s^2_A	t	df	P (>t)	F	df	P (>F)
Time	−.719	6.42×10^{-7}	.85703	1.45×10^{-6}	.0320	160	.487	2.25	80	.000177[a]
Linear	−.697	1.53×10^{-9}	.847	2.37×10^{-9}	.00151	160	.499	1.55	80	.0260[a]
Thickness	−.773	.000789	.850	.00142	1.72	160	.0434[a]	1.81	80	.00443[a]

Note. $\Delta X_W/\Delta T_W$ and s^2_W refer to mean rate and variance estimated within the ancestral series; $\Delta X_A/\Delta T_A$ and s^2_A refer to mean rate and variance estimated among the ancestral and descendant series; t-test performed on mean rates; F-test performed on rate variance.
[a] Implies rejection of the null hypothesis of gradualism at the 0.05 level or below; for details, see "The Charlesworth-Cheetham Test" in the appendix available in the online edition or from the *Journal of Geology* office.

0.05 level. The Charlesworth-Cheetham test is a very stringent test for punctuated equilibrium, and the mere possibility of Type I error caused by irregular subsampling underscores the importance of understanding depositional architecture and age model uncertainty.

One possible way to generate a microevolutionary pattern of punctuated equilibrium with this model is through a population bottleneck, which can cause a rapid shift in the phenotypic mean (fig. 4). A bottleneck can be induced by a constriction of suitable habitat, for instance during a regression. By assigning a narrow grain size preference, the species modeled in figure 7 is forced through a prolonged bottleneck (or series of bottlenecks). As shown in figure 7*E*, the stratophenetic series dutifully capture the periods of relative stasis but miss the phenotypic transition itself, as expected by punctuated equilibrium theory. Figure 7*E* also illustrates the unfortunate "bottleneck problem": a population bottleneck and associated phenotypic shift caused by this kind of reduction in available habitat (e.g., regression beyond the shelf break) will coincide with a major depositional hiatus, which is exactly where a phenotypic shift may appear in the stratophenetic series as an artifact of incompleteness. Teasing apart these confounding factors will require incorporation of several types of data and more sophisticated analytical tools.

A Cautionary Note on Ecophenotypy. Ecophenotypy is a complex topic that cannot be given sufficient treatment in this article, but some aspects of ecophenotypic variation relevant to the results of the above experiments deserve mention. Although not implemented in the experiments described above, the model can include ecophenotypic variation in the form of a linear reaction norm for a continuous macroenvironmental variable (see "The Price Equation" in the appendix). The imposition of a simple ecophenotypic gradient can cause differences in observed stratophenetic pattern among locations and apparent stasis due to upsection environmental fluctuations. Ecophenotypy also increases overall phenotypic variance and can cause spurious directionality as a result of facies trends.

However, these first-order descriptions of the effects of ecophenotypy, which can be made without recourse to numerical modeling, belie the true complexity of the issue. Environmental fluctuations can set up fluctuating selection pressures that will be reflected phenotypically either as evolutionary responses in heritable traits (Grant and Grant 1995) or as adaptive phenotypic plasticity, which can have different types of genetic control and can itself be the target of selection (Via et al. 1995; Schlichting and Pigliucci 1998). Without experimental evidence from living representatives, a distinction between environmentally driven selection responses on the one hand and phenotypic plasticity on the other may be intractable. A fair treatment of this topic thus goes beyond the scope of this article. The point to emphasize here is that, notwithstanding the challenges, a model of stratophenetic data must be able to account for fluctuating selection as well as ecophenotypy, and if relevant data are collected, inferring effects of underlying factors becomes a matter of estimating probabilities conditioned on the data.

Discussion

A numerical model of microevolution in the fossil record provides valuable lessons on the nature of stratophenetic data, shedding light on both empirical and analytical aspects. Some of these lessons are relevant to stratigraphic studies of other geological and geochemical variables. Analytically, series length is a major concern. Statistical tests like the scaled maximum test, which is based on limiting properties of infinite series, can give misleading results for short series. Similarly, the runs test assumes that the distribution of runs is normal, which may be violated in series with a small number of runs (caveat: for very short series, the normal approximation can be avoided by using the exact probability equations). Hurst exponent estimates for series of <500 data points are unreliable indi-

cators of the generating process and are sensitive to irregularities (jumps and spikes) in the series (Katsev and L'Hereux 2003). Thus, the confidence intervals for random walks of <50 data points start to approach the theoretically expected limits for the Hurst exponent, rendering the test impotent (caveat: if the time series is sampled regularly, methods like rescaled range analysis can give narrower confidence intervals). Furthermore, if a time series is subsampled irregularly (e.g., randomly), then the accuracy of sample age estimates becomes an important determinant of the Hurst exponent value. For example, if time is unknown and a randomly sampled series is analyzed using stratigraphic position (or sample numbers; e.g., Roopnarine et al. 1999), then the series may be effectively randomized, and the Hurst exponent follows suit.

Problems regarding the statistical behavior of random-walk tests applied to fossil data have been recognized and discussed at length by others (see Roopnarine et al. 1999; Sheets and Mitchell 2001). However, little attention has been paid to analytical stasis, the tendency toward Type I error in favor of stasis, which has potentially important implications for the empirical determination of relative frequency of modes of evolution in the fossil record. Analytical stasis in stratophenetic series can be caused by "static noise" (antipersistence, fluctuations) from several sources: (1) inaccurate estimation of population values, (2) sampling at irregular time intervals, (3) inaccurate estimation of sample ages, (4) geographic/spatial phenotypic variation, (5) ecophenotypy, and (6) migration. The good news is that the fossil record can yield information on most of these sources of noise. Like all scientists, paleontologists will never be in the business of analyzing noise-free data, but we can vastly improve our methods of inference by incorporating quantitative measures of uncertainty in the data available to us.

Stasis is a particularly important aspect of stratophenetic series to understand because of the importance placed on it in current evolutionary paleobiology (Gould and Eldredge 1993; Erwin and Anstey 1995; Jackson and Cheetham 1999; Jablonski 2000). It is clear, however, that statistical tests designed to detect stasis in the form of oscillations through time are sensitive to sampling error associated with sample size (which will make nonstatic data appear static) and temporally irregular sampling (which can make static data appear nonstatic or vice versa). Furthermore, stasis caused by stabilizing selection (a reduction in variance) or other genuine temporal changes in phenotypic variance will not be detected by tests focused on patterns in a single value. A more biologically plausible mechanism for long-term stasis may be fluctuations in selection direction and intensity. To evaluate the role of different mechanisms for causing an observed pattern of stasis, given the range of possibilities (sampling, stabilizing selection, fluctuating selection), we need not only a parameterization of these mechanisms but also a framework for model selection to determine the model structure best supported by available data.

Studies of microevolutionary rate in modern populations, where the vagaries of preservation and stratigraphy do not apply, are also susceptible to analytical stasis from inaccurate estimation of population parameters (see discussions in Kinnison and Hendry 2001; Merilä et al. 2001). Clearly, geographic (spatial/environmental) coverage is an important control on population parameter estimates, which raises the bar for empirical sampling protocols (Gould and Eldredge 1977; Jablonski 2000). Paleontological studies in areas with sufficient available outcrop can take advantage of the central-limit theorem and use the grand mean to improve accuracy by collecting many replicate samples along laterally extensive beds (Bennington 2003). Samples from sedimentary cores have to rely on microfossils and need to estimate phenotypic variability from modern representatives, when available (e.g., Malmgren et al. 1983), or, if extinct, from synchronous "event beds" found in multiple cores.

The spurious punctuation pattern in figure 6*E* and table 3 is just a single realization but might still confer some generality. Creation of nonrandom patterns as an artifact of the basin fill is a cause for concern, given the following points:

1. The distal series is the most complete of the stratophenetic series. We can measure completeness simply as the proportion of time steps represented in the stratophenetic series (0.107 for the most distal series, 0.007 for the most proximal series) or as a metric based on depositional completeness (the proportion of time steps represented in the final deposits), which yields 0.72 for the distal section and 0.24 for the proximal section. By comparison, Cheetham (1986) reports an expected depositional completeness of 0.63 at approximately 0.2-m.yr. resolution over 5 m.yr. In other words, although overall depositional completeness is essential, we also need to explicitly consider the unsteadiness of deposition upsection and across the study area when analyzing and interpreting stratophenetic patterns in the shallow marine fossil record. The necessary stratigraphic information (e.g., high-resolution sequence stratigraphic analysis) that might allow an evaluation of temporal

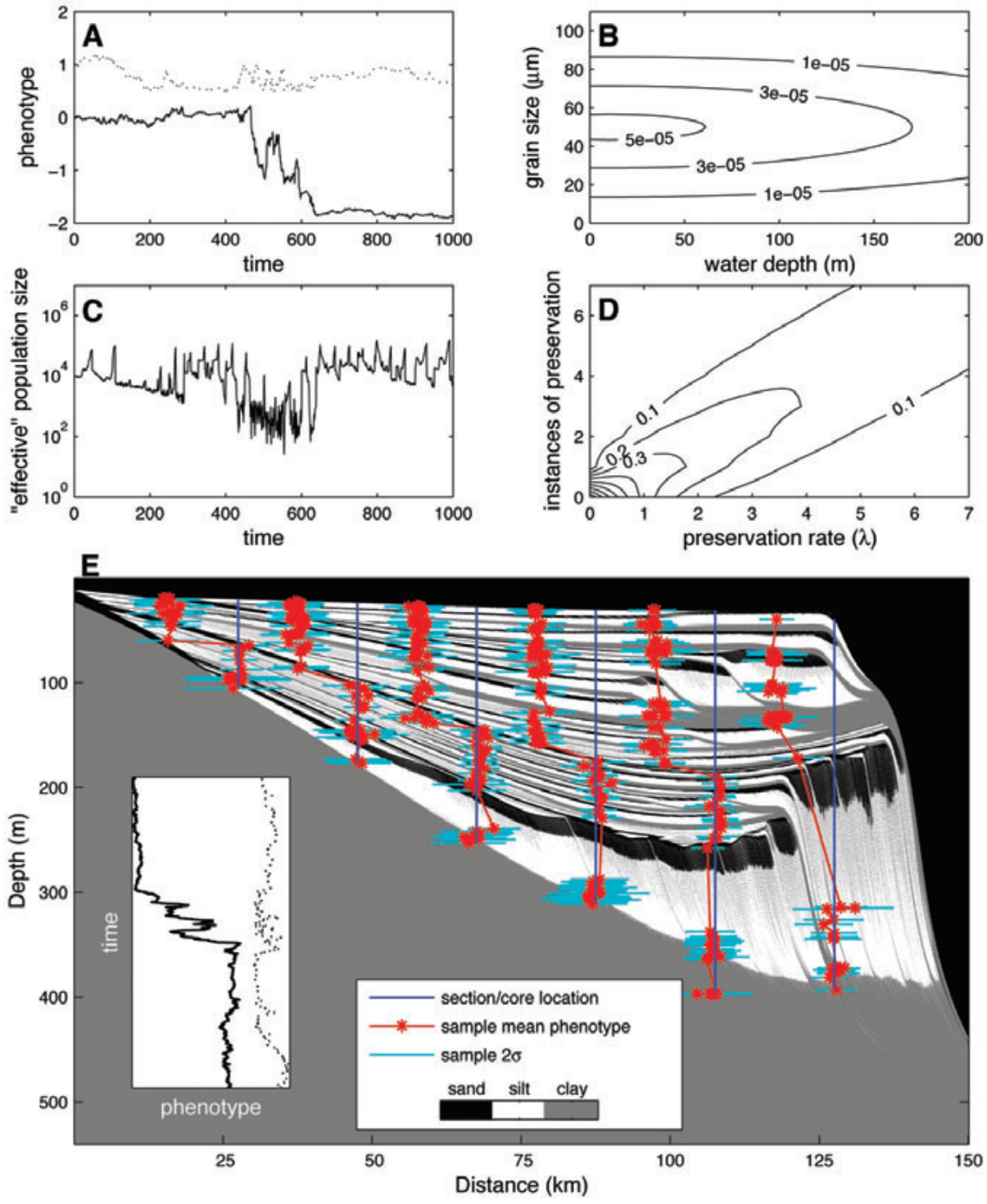

Figure 7. Numerical experiment with a population bottleneck. *A*, Phenotypic evolution based on parameters $\beta_{W,\phi} = 0$, $\beta_{W,(\phi-\bar{\phi})^2} = 0$, and N from *C*. Solid line is for phenotypic mean; dotted line is for variance. *B*, Gaussian density based on preferred depth = 10 m, depth tolerance = 150 m, preferred grain size = 50 μm, grain size tolerance = 20 μm. *C*, Population size (equal to "effective" population size) through time, based on the model in *B*. Note very small population size corresponding to the rapid shift in phenotypic mean in *A*. *D*, Poisson density function used to determine preservation based on λ as a function of the abundance in each model bin. *E*, Stratophenetic series plotted against synthetic basin fill from figure 2. Sample means (*red*) and error bars (*cyan*) are scaled to make the stratophenetic

heterogeneity on independent lithological grounds is often unavailable, making it difficult to assess the effects of depositional architecture on empirical patterns reported in the literature.

2. The distal part of the synthetic basin fill used in this numerical experiment (see fig. 2) is characterized not by unusually large stratigraphic gaps in terms of the absolute time intervals represented by nondeposition or erosion but rather by high temporal completeness of the deposits between gaps, preserving populations that are closely spaced in time (i.e., making the gaps large only in a relative sense). Furthermore, on the timescales considered, much of this depositional heterogeneity would be below biostratigraphic resolution. Analyzing multiple stratophenetic series from different localities is an essential first step toward building a more complete composite picture (Gould and Eldredge 1977), but even with such data, it would be difficult to make sense of patterns without sufficient knowledge of basin stratigraphy.

3. Even with good biostratigraphic control and radiometric dating, error bars on age estimates of directly dated samples tend to be large, and many sample ages have to be interpolated between dated points. The uncertainties in these age estimates need to be quantified and incorporated in the analysis of stratophenetic series, and any available high-resolution lithological, paleoecological, and geochemical information on subtle temporal and environmental heterogeneity should be used to evaluate the magnitude and variance of this uncertainty.

4. Random subsampling (based on uniformly distributed pseudorandom numbers) of the microevolutionary series cannot account for the artifacts of irregular sampling caused by the basin fill. Even statistical tests that are robust to series length will have little power to resolve the underlying pattern unless they can somehow digest contextual information about the stratophenetic series, such as uncertainty in the temporal spacing of samples.

Stratigraphic overprint is a perennial problem for paleontologists, and many authors have emphasized the importance of stratigraphy to the evolutionary interpretation of patterns in the fossil record (e.g., Brinkmann 1929; Newell 1956; Jablonski 1980; MacLeod 1991; Holland and Patzkowsky 1999; Kidwell and Holland 2002). Nevertheless, many empirical studies analyze stratophenetic series with samples plotted against stratigraphic position (e.g., Gingerich 1976*b*; Raup and Crick 1981; Reyment 1982, 1985; Kelley 1983; Sheldon 1987; McCormick and Fortey 2002), and this is a particular concern for data from shallow marine and terrestrial depositional settings, where sedimentation rates and environments are highly variable. Several studies include absolute age constraints (e.g., Williamson 1981; Cheetham 1986; Geary 1987; Stanley and Yang 1987), and some include comprehensive age models, especially for deep-sea cores (e.g., Kellogg and Hays 1975; Malmgren et al. 1983; Whatley 1985; Wei 1994; Kucera and Malmgren 1998). However, none of these studies includes error bars on the ages of samples, although, presumably, the age models have considerable uncertainty (with the possible exception of varved lake deposits; Bell et al. 1985), which will propagate into estimates of evolutionary rate (MacLeod 1991). In shallow marine deposits, upsection changes in environment (which can be subtle) and in sedimentation rate, combined with habitat tracking and ecophenotypic variation, compromise the straightforward evolutionary interpretation of stratophenetic series (e.g., Reyment 1985; Smith and Paul 1985; Abe et al. 1988; McGhee et al. 1991; Daley 1999).

Many sources of data are available to expand the information content of stratophenetic series. In addition to more comprehensive shape quantification beyond single characters (Cheetham 1987), this can include sedimentological and stratigraphic information (grain size variation, facies analysis, and sequence stratigraphic interpretations), paleontological information (fossil density, faunal associations and biofacies, biostratigraphy), and geophysical/geochemical information (wireline logs, radiometric dates, $\delta^{18}O$, $\delta^{13}C$, and backstripping). Such data provide an essential context, generally in the form of estimates or proxies with variable degrees of uncertainty. Although empirical sampling protocols always leave room for improvement in terms of geographic coverage and temporal resolution, the development of new analytical protocols is also required.

Because we cannot measure population variation

patterns easily visible and have to be projected back onto the blue line ("core") for the samples to be in original stratigraphic position. *Inset*, microevolutionary series from *A*. At most localitites, the shift in phenotypic mean coincides with a stratigraphic hiatus, which is also the expected locus of artificial jumps.

or microevolutionary rates directly, we need to place observed morphological changes in their environmental and temporal context through methods of estimation conditioned on the data. Given an observed stratophenetic series and associated time-environment data, our knowledge of the system, and a model mapping evolutionary to stratophenetic patterns, what is the most plausible evolutionary pattern supported by the data? The many parameters and the stochasticity of the model described in this article make the number of possible outcomes too vast to allow an exhaustive exploration of the parameter space in a forward sense, but methods exist for sampling the important regions of this space. Rather than seeking general "laws" that govern stratophenetic series, this kind of model is meant to provide a tool for making inferences from empirical data. Work is currently in progress on applying Monte Carlo algorithms to stratophenetic inverse problems of this kind (Hannisdal 2003), which uses a similar forward model strategy as a template for parameter estimation by defining a probability distribution of models conditioned on observed data. Given the uncertainty of both data and model, how much information can be retrieved from stratophenetic series is still an open question of fundamental importance to evolutionary paleobiology.

Conclusions

The following three aspects of sampling error affect stratophenetic series: (*a*) short series length severely limits the applicability of statistical techniques from time series analysis (e.g., runs test, scaled maximum test, Hurst exponent); (*b*) small sample size can cause analytical stasis: the inaccuracy of sample mean values creates fluctuations in observed stratophenetic patterns, which will be identified as stasis, regardless of the underlying microevolutionary trajectory; and (*c*) sampling at irregular time intervals can cause nonstatic patterns to appear static or vice versa, a problem that is augmented with increasing age model error.

Depositional architecture exerts a fundamental control on stratophenetic series by determining the temporal spacing and environmental heterogeneity of successive samples. Even for relatively complete stratigraphic sections, the unsteadiness of deposition can generate spurious patterns in observed stratophenetic series, and the relative size of gaps may thus be more important than the absolute size of gaps. Ecophenotypic variation may also conspire with upsection environmental changes to confound evolutionary interpretations of observed patterns. Furthermore, a population bottleneck and rapid phenotypic shift caused by shelf exposure during regression will coincide with a depositional hiatus, which is the expected locus of artificial phenotypic jumps.

Fortunately, the fossil record does contain information that can be used to assess the importance of these potential confounding factors. Thus, to meet the challenges facing the analysis of stratophenetic series, many of which have been acknowledged qualitatively for more than a century, paleontologists need to take full advantage of available sources of information in the sedimentary record. Evolutionary inferences should be based on a joint analysis of evidence internal to the series (e.g., statistical moments) and external time-environment data. Quantifying the uncertainty of models and data will then find its place at the top of the paleontologists' agenda.

ACKNOWLEDGMENTS

Many thanks to E. Hutton, I. Overeem, and J. Syvitski at the Institute of Arctic and Alpine Research in Boulder, Colorado, for generous help with SedFlux. I am grateful to M. Foote, S. Holland, D. Jablonski, P. Wagner, and particularly S. Kidwell for discussion and comments on the manuscript. A. Miller and two anonymous reviewers provided very insightful reviews. Funding was provided by the Research Council of Norway, the University of Chicago Department of the Geophysical Sciences, and the American Chemical Society Petroleum Research Fund grant 41014-AC8 to S. Kidwell and M. Foote.

REFERENCES CITED

Abe, K.; Reyment, R.; Bookstein, F.; Honigstein, A.; Almogi-Labin, A.; Rosenfeld, A.; and Hermelin, O. 1988. Microevolution in two species of ostracods from the Santonian (Cretaceous) of Israel. Hist. Biol. 1:303–322.

Arnold, S. J.; Pfrender, M. E.; and Jones, A. G. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 112–113:9–32.

Bell, M. A.; Baumgartner, J. V.; and Olson, E. C. 1985. Patterns of temporal change in single morphological

characters of a Miocene stickleback fish. Paleobiology 11:258–271.
Bennington, J. B. 2003. Transcending patchiness in the comparative analysis of paleocommunities: a test case from the Upper Cretaceous of New Jersey. Palaios 18: 22–33.
Boogart, P. W.; Van Balen, R.; Kasse, C.; and Vandenberghe, J. 2003. Process-based modelling of fluvial system response to rapid climate change. I. Model formulation and generic applications. Q. Sci. Rev. 22: 2077–2095.
Bookstein, F. L. 1987. Random walks and the existence of evolutionary rates. Paleobiology 13:446–464.
Brinkmann, R. 1929. Statistisch-biostratigraphische Untersuchungen an mitteljurassischen Ammoniten über Artbegriff und Stammesentwicklung. Abh. Ges. Wiss. Gött. 13:1–249.
Brown, J. H. 1995. Macroecology. Chicago, University of Chicago Press, 269 p.
Bush, A. M.; Powell, M.; Arnold, W.; Bert, T.; and Daley, G. 2002. Time-averaging, evolution, and morphologic variation. Paleobiology 28:9–25.
Charlesworth, B. 1984. Some quantitative methods for studying evolutionary patterns in single characters. Paleobiology 10:308–318.
Charlesworth, B.; Lande, R.; and Slatkin, M. 1982. A neo-Darwinian commentary on macroevolution. Evolution 36:474–498.
Cheetham, A. H. 1986. Tempo of evolution in a Neogene bryozoan: rates of morphological change within and across species boundaries. Paleobiology 12:190–202.
———. 1987. Tempo of evolution in a Neogene bryozoan: are trends in single morphological characters misleading? Paleobiology 13:286–296.
Daley, G. M. 1999. Environmentally controlled variation in shell size of *Ambonychia* (Mollusca: Bivalvia) in the type Cincinnatian (Upper Ordovician). Palaios 14: 520–529.
De Rijk, S.; Jorissen, F.; Rohling, E.; and Troelstra, S. 2000. Organic flux control on bathymetric zonation of Mediterranean benthic foraminifera. Mar. Micropaleontol. 40:151–166.
Eldredge, N., and Gould, S. J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. *In* Schopf, T. J. M., ed. Models in paleobiology. New York, Freeman, p. 82–115.
Erwin, D. H., and Anstey, R. L. 1995. Speciation in the fossil record. *In* Erwin, D. H., and Anstey, R. L., eds. New approaches to speciation in the fossil record. New York, Columbia University Press, p. 11–38.
Frank, S. A. 1997. The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. Evolution 51:1712–1729.
Geary, D. H. 1987. Evolutionary tempo and mode in a sequence of the Upper Cretaceous bivalve *Pleurocardia*. Paleobiology 13:140–151.
Gingerich, P. D. 1976*a*. Cranial anatomy and evolution of North American Plesiadapidae (Mammalia, Primates). Univ. Mich. Pap. Paleontol. 15:1–140.
———. 1976*b*. Paleontology and phylogeny: patterns of evolution at the species level in early Tertiary mammals. Am. J. Sci. 276:1–28.
———. 2001. Rates of evolution on the time scale of the evolutionary process. Genetica 112–113:127–144.
Girard, C.; Renaud, S.; and Korn, D. 2004. Step-wise morphological trends in fluctuating environments: evidence in the Late Devonian conodont genus *Palmatolepis*. Geobios 37:404–415.
Gould, S. J., and Eldredge, N. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology 3:115–151.
———. 1993. Punctuated equilibrium comes of age. Nature 366:223–227.
Grant, P. R., and Grant, B. R. 1995. Predicting microevolutionary responses to directional selection on heritable variation. Evolution 49:241–251.
Gray, J. 1981. The ecology of marine sediments. Cambridge, Cambridge University Press, 185 p.
Hannisdal, B. 2003. Inferring evolutionary patterns in the fossil record using Bayesian inversion: an application to synthetic stratophenetic data. Geol. Soc. Am. Abstr. Program 35:165.
Hendry, A. P., and Kinnison, M. T. 1999. Perspective: the pace of modern life: measuring rates of contemporary microevolution. Evolution 53:1637–1653.
Holland, S. M. 1995. The stratigraphic distribution of fossils. Paleobiology 21:92–109.
———. 2000. The quality of the fossil record: a sequence stratigraphic perspective. *In* Erwin, D. H., and Wing, S. L., eds. Deep time: *Paleobiology*'s perspective. Paleobiology 26(suppl.):148–168.
Holland, S. M., and Patzkowsky, M. E. 1999. Models for simulating the fossil record. Geology 27:491–494.
Hunt, G. 2004*a*. Phenotypic variance inflation in fossil samples: an empirical assessment. Paleobiology 30: 487–506.
———. 2004*b*. Phenotypic variation in fossil samples: modeling the consequences of time-averaging. Paleobiology 30:426–443.
Hutton, E. W. H., and Syvitski, J. P. M. 2004. Advances in the numerical modeling of sediment failure during the development of a continental margin. Mar. Geol. 203:367–380.
Jablonski, D. 1980. Apparent versus real biotic effects of transgressions and regressions. Paleobiology 6:397–407.
———. 2000. Micro- and macroevolution: scale and hierarchy in evolutionary biology and paleobiology. *In* Erwin, D. H., and Wing, S. L., eds. Deep time: *Paleobiology*'s perspective. Paleobiology 26(suppl.):15–52.
Jackson, J. B. C., and Cheetham, A. H. 1999. Tempo and mode of speciation in the sea. Trends Ecol. Evol. 14: 72–77.
Katsev, S., and L'Hereux, I. L. 2003. Are Hurst exponents estimated from short or irregular time series meaningful? Comput. Geosci. 29:1085–1089.
Kelley, P. H. 1983. Evolutionary patterns of 8 Chesapeake Group mollusks: evidence for the model of punctuated equilibria. J. Paleontol. 57:581–598.
Kellogg, D. E., and Hays, J. D. 1975. Microevolutionary

patterns in Late Cenozoic Radiolaria. Paleobiology 1: 150–160.
Kidwell, S. M., and Holland, S. M. 2002. The quality of the fossil record: implications for evolutionary analyses. Annu. Rev. Ecol. Syst. 33:561–588.
Kinnison, M. T., and Hendry, A. P. 2001. The pace of modern life. II. From rates of contemporary microevolution to pattern and process. Genetica 112–113: 145–164.
Kowalewski, M., and Bambach, R. K. 2003. The limits of paleontological resolution. *In* Harries, P. J., ed. High resolution approaches in stratigraphic paleontology. New York, Plenum, p. 1–48.
Kucera, M., and Malmgren, B. A. 1998. Differences between evolution of mean form and evolution of new morphotypes: an example from Late Cretaceous planktonic foraminifera. Paleobiology 24:49–63.
Lenihan, H. S., and Micheli, F. 2001. Soft-sediment communities. *In* Bertness, M. D.; Gaines, S. D.; and Hay, M. E., eds. Marine community ecology. Sunderland, MA, Sinauer, p. 253–287.
Levinton, J. S. 1982. Marine ecology. Englewood Cliffs, NJ, Prentice-Hall, 526 p.
MacLeod, N. 1991. Punctuated anagenesis and the importance of stratigraphy to paleobiology. Paleobiology 17:167–188.
Malmgren, B. A.; Berggren, W. A.; and Lohmann, G. P. 1983. Evidence for punctuated gradualism in the Late Neogene *Globorotalia tumida* lineage of planktonic foraminifera. Paleobiology 9:377–389.
McCormick, T., and Fortey, R. A. 2002. The Ordovician trilobite *Carolinites*, a test case for microevolution in a macrofossil lineage. Palaeontology 45:229–257.
McGhee, G. J.; Bayer, U.; and Seilacher, A. 1991. Biological and evolutionary responses to transgressive-regressive cycles. *In* Ricken, W., and Seilacher, A., eds. Cycles and events in stratigraphy. Berlin, Springer, p. 696–708.
Merilä, J.; Sheldon, B. C.; and Kruuk, L. 2001. Explaining stasis: microevolutionary studies in natural populations. Genetica 112–113:199–222.
Morehead, M. D.; Syvitski, J. P. M.; and Hutton, E. W. H. 2001. The link between abrupt climate change and basin stratigraphy: a numerical approach. Global Planet. Change 28:107–127.
Murray, J. W. 1991. Ecology and paleoecology of benthic foraminifera. New York, Wiley, 397 p.
Newell, N. D. 1956. Catastrophism and the fossil record. Evolution 10:97–101.
O'Grady, D. B., and Syvitski, J. P. M. 2001. Predicting profile geometry of continental slopes with a multiprocess sedimentation model. *In* Merriam, D. F., and Davis, J. C., eds. Geological modeling and simulation: sedimentary systems. New York, Kluwer Academic, p. 99–117.
Paola, C. 2000. Quantitative models of sedimentary basin filling. Sedimentology 47:121–178.
Parsons, T. R.; Takahashi, M.; and Hargrave, B. 1984. Biological oceanographic processes. 3rd ed. Oxford, Pergamon, 320 p.
Perlmutter, M., and Plotnick, R. E. 2003. Hemispheric asymmetry of the stratigraphic record: conceptual proof of unipolar glaciation. *In* Cecil, C., and Edgar, N., eds. Climate controls on stratigraphy. Tulsa, OK, SEPM Spec. Publ. 77, p. 51–68.
Plotnick, R. E. 1996. The ecological play and the geological theater. Palaios 11:1–2.
Price, G. R. 1970. Selection and covariance. Nature 227: 520–521.
Raup, D. M., and Crick, R. E. 1981. Evolution of single characters in the Jurassic ammonite *Kosmoceras*. Paleobiology 7:200–215.
Reyment, R. A. 1982. Phenotypic evolution in a Cretaceous foraminifer. Evolution 36:1182–1199.
———. 1985. Phenotypic evolution in a lineage of the Eocene ostracod *Echinocythereis*. Paleobiology 11: 174–194.
Reznick, D. N.; Shaw, F. H.; Rodd, F. H.; and Shaw, R. G. 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). Science 275:1934–1937.
Rice, S. H. 2004. Evolutionary theory. Sunderland, MA, Sinauer, 370 p.
Roopnarine, P. D. 2001. The description and classification of evolutionary mode: a computational approach. Paleobiology 27:446–465.
Roopnarine, P. D.; Byars, G.; and Fitzgerald, P. 1999. Anagenetic evolution, stratophenetic patterns, and random walk models. Paleobiology 25:41–57.
Rowe, A. 1899. Analysis of the genus *Micraster*, as determined by rigid zonal collection from the zone of *Rhynchonella cuvieri* to that of *Micraster coranguinum*. Q. J. Geol. Soc. Lond. 55:494–546.
Sadler, P. M. 1981. Sediment accumulation rates and the completeness of stratigraphic sections. J. Geol. 89: 569–584.
Schlichting, C. D., and Pigliucci, M. 1998. Phenotypic evolution: a reaction norm perspective. Sunderland, MA, Sinauer, 387 p.
Schluter, D. 2000. The ecology of adaptive radiation. Oxford, Oxford University Press, 288 p.
Shafer, S. L.; Bartlein, P. J.; and Thompson, R. S. 2001. Potential changes in the distributions of western North America tree and shrub taxa under future climate scenarios. Ecosystems 4:200–215.
Sheets, D. H., and Mitchell, C. E. 2001. Why the null matters: statistical tests, random walks and evolution. Genetica 112–113:105–125.
Sheldon, P. R. 1987. Parallel gradualistic evolution of Ordovician trilobites. Nature 330:561–563.
Simpson, G. G. 1944. Tempo and mode in evolution. New York, Columbia University Press, 237 p.
Sisson, J. D.; Shimeta, J.; Zimmer, C. A.; and Traykovski, P. 2002. Mapping epibenthic assemblages and their relations to sedimentary features in shallow-water, high-energy environments. Cont. Shelf Res. 22:565–583.
Smith, A. B., and Paul, C. R. 1985. Variation in the irregular echinoid *Discoides* during the Early Cenomanian. *In* Cope, J. C. W., and Skelton, P. W., eds.

Evolutionary case histories from the fossil record. Spec. Pap. Palaeontol. 33:29–37.
Stanley, S. M. 1979. Macroevolution. San Francisco, W. H. Freeman, 332 p.
Stanley, S. M., and Yang, X. 1987. Approximate evolutionary stasis for bivalve morphology over millions of years: a multivariate, multilineage study. Paleobiology 13:113–139.
Syvitski, J. P. M., and Bahr, D. B. 2001. Numerical models of marine sediment transport and deposition. Comput. Geosci. 27:617–618.
Syvitski, J. P. M., and Hutton, E. W. H. 2001. 2D SEDFLUX 1.0C: an advanced process-response numerical model for the fill of marine sedimentary basins. Comput. Geosci. 27:731–753.
Van Hoey, G.; Degraer, S.; and Vincx, M. 2004. Macrobenthic community structure of soft-bottom sediments at the Belgian continental shelf. Estuar. Coast. Shelf Sci. 59:599–613.
Via, S.; Gomulkiewicz, R.; Dejong, G.; Scheiner, S.; Schlichting, C.; and Vantienderen, P. 1995. Adaptive phenotypic plasticity: consensus and controversy. Trends Ecol. Evol. 10:212–217.
Wackerly, D. D.; Mendenhall, W., III; and Scheaffer, R. L. 2002. Mathematical statistics with applications. 6th ed. Pacific Grove, CA, Duxbury, 853 p.
Wei, K.-Y. 1994. Stratophenetic tracing of phylogeny using SIMCA pattern recognition technique: a case study of the late Neogene planktic foraminifera *Globoconella* clade. Paleobiology 20:52–65.
Whatley, R. C. 1985. Evolution of the ostracods *Bradleya* and *Poseidonamicus* in the deep-sea Cainozoic of the south-west Pacific. *In* Cope, J. C. W., and Skelton, P. W., eds. Evolutionary case histories from the fossil record. Spec. Pap. Palaeontol. 33:103–116.
Whittaker, R. H. 1970. Communities and ecosystems. New York, Macmillan, 158 p.
Williamson, P. G. 1981. Palaeontological documentation of speciation in Cenozoic molluscs from Turkana Basin. Nature 293:437–443.

Appendix from B. Hannisdal, "Phenotypic Evolution in the Fossil Record: Numerical Experiments"

(J. Geol., vol. 114, no. 2, p. 133)

Details of Numerical Experiments

The Price Equation

This article follows the notation of Rice (2004), using the following quantities: N = population size, ϕ_i = phenotype of individual i, $\overline{\phi}$ = population mean phenotype, $\delta_{i,j}$ = difference between the phenotype of i and that of its jth descendant, $\overline{\delta_i}$ = difference between ϕ_i and the mean value of i's descendants, W_i = number of descendants of i, and $\overline{W}$ = mean number of descendants per individual. If we express the phenotype of descendant j of individual i as $\phi_i + \delta_{i,j}$, the mean phenotype of descendants is given by

$$\overline{\phi'} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{W_i}(\phi_i + \delta_{i,j})}{\sum_{i=1}^{N} W_i}. \tag{A1}$$

From the definition of the arithmetic mean, we have

$$\sum_{j=1}^{W_i} \phi_i = W_i\phi_i,$$

$$\sum_{j=1}^{W_i} \delta_{i,j} = W_i\overline{\delta}_i,$$

$$\sum_{i=1}^{N} W_i = N\overline{W},$$

which allows rewriting equation (A1) as

$$\begin{aligned}\overline{\phi'} &= \frac{1}{N\overline{W}}\left(\sum_{i=1}^{N} W_i\phi_i + \sum_{i=1}^{N} W_i\overline{\delta}_i\right)\\ &= \frac{1}{\overline{W}}[\mathrm{E}(W\phi) + \mathrm{E}(W\overline{\delta})].\end{aligned} \tag{A2}$$

The identity $\mathrm{Cov}(x, y) = \mathrm{E}(xy) - \overline{x}\,\overline{y}$ can then be used to rewrite equation (A2):

$$\begin{aligned}\overline{\phi'} &= \frac{1}{\overline{W}}[\mathrm{Cov}(W, \phi) + \overline{W}\,\overline{\phi} + \mathrm{E}(W\overline{\delta})]\\ &= \frac{1}{\overline{W}}[\mathrm{Cov}(W, \phi) + \mathrm{E}(W\overline{\delta})] + \overline{\phi},\end{aligned} \tag{A3}$$

which yields the Price equation in its standard form:

$$\Delta\overline{\phi} = \frac{1}{\overline{W}}[\mathrm{Cov}(W, \phi) + E(W\overline{\delta})]. \tag{A4}$$

For more details on equation (A4), see work by Price (1970), Frank (1997), and Rice (2004).

Transmission effects normally included in the $\delta_{i,j}$ term (important for modeling evolution on a generational timescale) are here replaced by the effect of environmental (e.g., water depth and/or substrate properties) variability on phenotype. A simple scenario is one in which the relationship between phenotype and environment is described as a linear regression with random noise (i.e., a linear reaction norm for a continuous environmental variable):

$$\delta_{i,j} = \beta_{\phi,e}(e_j - e_i), \tag{A5}$$

where e_i and e_j are the environments of ancestor and descendant, respectively, and $\beta_{\phi,e}$ is the linear regression coefficient of phenotype on environment. If we assume that the norm of reaction is not itself under selection, then $\mathrm{Cov}(W, \bar{\delta}) = 0$, and we can use the fact that $\mathrm{Cov}(x, y) = \beta_{x,y}\mathrm{Var}(y)$ to rewrite equation (A4):

$$\Delta\bar{\phi} = \frac{1}{\bar{W}}\beta_{W,\phi}\mathrm{Var}(\phi) + \bar{\delta}, \tag{A6}$$

where

$$\bar{\delta} = \beta_{\phi,e}\Delta\bar{e}. \tag{A7}$$

An expression for the change in phenotypic variance is found by simply substituting $(\phi - \bar{\phi})^2$ for ϕ in equation (A6), and if we set $\beta_{(\phi-\bar{\phi})^2,e} = 0$ in equation (A7), we get the following expression for the change in phenotypic variance:

$$\Delta\mathrm{Var}(\phi) = \frac{1}{\bar{W}}\beta_{W,(\phi-\bar{\phi})^2}\mathrm{Var}[(\phi - \bar{\phi})^2]. \tag{A8}$$

The expected fitness distribution for selectively neutral variation, with no population growth or decline due to differential reproductive success, is a Poisson with mean $\bar{W} = 1$ (and thus a variance of 1), which can be shown to yield the following simple expression for the variance in phenotypic change (Rice 2004):

$$\mathrm{Var}(\Delta\bar{\phi}) = \frac{\mathrm{Var}(\phi)}{N}. \tag{A9}$$

Drift is thus caused by finite sampling of descendants from the ancestral distribution.

The Runs Test

A run refers to a sequence of consecutive positive or negative steps in a series (first differences). Raup and Crick (1981) base their test on the equations for the exact probability of a given number of runs. However, for longer series (>50 data points), the factorial products make this equation computationally unwieldy, and an approximation is used. Siegel and Castellan (1988) give a Z score based on the expected mean $\bar{R}$ and variance s^2 of the number of runs as a function of the total number of positive steps N_1 and negative steps N_2 in the series with observed number of runs R:

$$Z = \frac{R - \bar{R}}{s}, \tag{A10}$$

where

$$\bar{R} = 1 + \frac{2N_1N_2}{N_1 + N_2}, \tag{A11}$$

and

$$s = \left[\frac{2N_1N_2(2N_1N_2 - N_1 - N_2)}{(N_1 + N_2)^2(N_1 + N_2 - 1)}\right]^{1/2}. \tag{A12}$$

The Z score is then compared to a standard normal table such that an absolute Z value >1.96 leads to rejection of a random-walk null hypothesis at the 0.05 significance level. This test thus assumes that the number of runs is normally distributed, which makes it suspect for short series with a small number of runs.

The Scaled Maximum Test

For a symmetric random walk where the step size has finite variance σ^2, the theorem of the scaled maximum basically compares the maximum position ($\max |S_k|$), where $1 < k < N$, of an observed series of length N to the mean maximum position expected for a random walk $\sigma\sqrt{N}$ (the standard deviation of the final step). Although it strictly applies only as N goes to infinity, it can be used to test whether an observed series is consistent with a random walk (Bookstein 1987, 1988) using the range statistic x:

$$x = \frac{\max |S_k|}{\sigma\sqrt{N}}, \tag{A13}$$

where $\sigma\sqrt{N}$ can be estimated as

$$\sigma\sqrt{N} \approx \left[\sum_i (S_{n_i} - S_{n_{i-1}})^2\right]^{1/2}. \tag{A14}$$

Bookstein (1987) gives the following values as a confidence interval for the x statistic: $x > 2.25$ leads to rejection of the random walk in favor of anagenesis, whereas $x < 0.62$ leads to rejection of the random walk in favor of stasis.

The Hurst Exponent

If a time series $y(t)$ is a self-affine fractal, then $y(bt)$ is statistically equivalent to $b^H y(t)$, where H is the Hurst exponent (Turcotte 1997). In other words, the positional range of the series is proportional to the time interval over which it is measured. White noise (a series of normally distributed random numbers) has a theoretically expected Hurst exponent $H = 0$; Brownian noise (the cumulative sum of white noise, i.e., a random walk) is a self-affine fractal with expected $H = 0.5$, and a perfectly self-similar process would have an expected $H = 1$. The Hurst exponent can therefore be used to characterize stochastic processes and is commonly used to distinguish random from static (antipersistent) and directional (persistent) patterns in noisy data series.

For a time series collected at regular intervals, a measure of the Hurst exponent can be obtained by a simple method based on the so-called series width (Katsev and L'Hereux 2003). For a time interval t, the series width $w(t)$ is given by

$$w(t) = \langle(\langle(y - \langle y\rangle_t)^2\rangle_t)^{1/2}\rangle, \tag{A15}$$

where $\langle y\rangle_t$ is the mean over the interval t and the outer brackets indicate averaging over all possible intervals of length t, $(1 \le t \le N)$. When the function $w(t)$ is plotted on a log-log plot, the slope of a linear least squares fit is a measure of the Hurst exponent. Note, however, that a linear relationship is normally limited to intermediate values of t, and estimates of H should be based on fitting a line only to an intermediate portion covering 50%–80% of the range of t values.

Roopnarine et al. (1999) used the Hurst exponent to characterize evolutionary mode in stratophenetic series by calculating rescaled range, with sample numbers providing equal time intervals. However, for a stratophenetic series that is sampling the microevolutionary time series at irregular points in time, the average values (of range or series width) and standard deviations for each interval length have to be replaced by pairwise differences between samples for all possible intervals. A least squares slope is then fitted to all the points on a log-log scale. This is the approach taken by Roopnarine (2001), who also extended the analysis with a bootstrap significance test. Thus, for two samples i and j, the Hurst exponent is estimated by the least squares slope of the function

$$\ln(|y_i - y_j|) = H \ln(|t_i - t_j|), j < i \leq N, \tag{A16}$$

which is analogous to the log rate interval method of Gingerich (1993).

For short series, H estimates obtained using this method are extremely unreliable, and in order to establish a significance test criterion, this article compares the estimate of H from an observed series with a distribution of H estimates for 500 random-walk realizations of the same length as the observed series. An H value lower than the 2.5 percentile of the distribution implies rejection of the random walk in favor of stasis, and a value higher than the 97.5 percentile implies rejection of the random walk in favor of anagenesis. See work by Bassingthwaighte and Raymond (1994) and Wallis and Matalas (1969) for more discussion of the effects of small sample size on the Hurst exponent.

The Charlesworth-Cheetham Test

Charlesworth (1984) described methods for estimating phenotypic evolutionary rates and variance, which were extended by Cheetham (1986) to test the null hypothesis that the rate of change within a species is sufficient to account for the rate of change between species. The mean rate of change within the ancestor species (species 1) is given by

$$\frac{\Delta X_W}{\Delta T_W} = \frac{X_{1,N} - X_{1,0}}{T_{1,N} - T_{1,0}}, \tag{A17}$$

where $T_{1,0}$ and $T_{1,N}$ are the first and last occurrences and $X_{1,0}$ and $X_{1,N}$ are the corresponding phenotypic values. The rate variance is defined as

$$s_W^2 = \frac{1}{\Delta T_W} \sum_{i=0}^{N-1} \left[\Delta X_{1,i} - (T_{1,i+1} - T_{1,i}) \frac{\Delta X_W}{\Delta T_W} \right]^2. \tag{A18}$$

The mean rate of change across the boundary between ancestor species 1 and descendant species 2 is given by

$$\frac{\Delta X_A}{\Delta T_A} = \frac{X_{2,0} - X_{1,0}}{T_{2,0} - T_{1,0}}, \tag{A19}$$

where $T_{2,0}$ is the first occurrence of the descendant species and $X_{2,0}$ is the corresponding phenotypic value. The across-species rate variance is defined as

$$s_A^2 = \frac{1}{\Delta T_W} \sum_{i=0}^{N-1} \left[\Delta X_{1,i} - (T_{1,i+1} - T_{1,i}) \frac{\Delta X_A}{\Delta T_A} \right]^2. \tag{A20}$$

A difference in mean rates is evaluated using a t-test:

$$t = \frac{[(\Delta X_A/\Delta T_A) - (\Delta X_W/\Delta T_W)](N/2)^{1/2}}{(s_A^2 + s_W^2)/2}, \tag{A21}$$

with $2N - 2$ df. A difference in rate variance is evaluated using an F-test:

$$F = \frac{s_A^2}{s_W^2}, \tag{A22}$$

with $N - 1$ df in both numerator and denominator.

SedFlux Input Parameter Values

The following tables list the sediment properties (table A1) and main input parameter values (table A2) used to generate the SedFlux basin fill predictions described in the article. Not included here are the sea level curve and the initial bathymetric profile. The model version used was 2DSedFlux 1.5, which includes several additional processes that were not activated. Note that processes have their own internal time steps. See work by Syvitski and Hutton (2001) for details.

Table A1
Sediment Properties

	Grain				
Property	1	2	3	4	5
Size (μm)	800	300	70	10	1
Density (kg/m^3)	2625	2600	2550	2500	2450
Saturated density (kg/m^3)	1850	1800	1750	1700	1650
Minimum void ratio	.30	.20	.15	.10	.050
Plastic index	.10	.20	.30	.40	.50
Diffusion coeffecient	1.00	.75	.50	.25	.25
Removal rate (d^{-1})	50	16.8	9	3.2	2.4
Consolidation coeffecient (m^2/yr)	1.0×10^{-10}	1.0×10^{-10}	1.0×10^{-10}	1.0×10^{-10}	1.0×10^{-10}
Compaction coeffecient	3.68×10^{-8}	5.0×10^{-8}	7.0×10^{-8}	8.0×10^{-8}	3.68×10^{-7}

Note. Grain 1 is bedload; the others are suspended load.

Table A2
Main SedFlux Input Parameter Values

Parameter	Value
Global settings:	
Duration (yr)	10,000
Time step (yr)	10
Vertical resolution (m)	.2
Horizontal resolution (m)	50
Basin length (km)	300
Basin width (km)	50
Isostasy:	
Flexural rigidity (Nm)	1.6×10^{22}
Relaxation time (yr)	2500
River input:	
Bedload (kg/s)	250
Suspended load (kg/m^3)	3, 2, 2, 2
River velocity (m/s)	1.0
River width (m)	200
River depth (m)	5
River input (transgressive phases):	
Bedload (kg/s)	100
Suspended load (kg/m^3)	2, 1, 1, 1
River velocity (m/s)	1.0
River width (m)	200
River depth (m)	2
Bedload:	
Distance to dump bedload (m)	5000
Ratio flood plain/bedload rate	.07
Plume:	

Table A2 (*Continued*)

Parameter	Value
Background ocean concentration	0
Maximum plume width (km)	3
No. grid nodes, cross-shore	9
No. grid nodes, river mouth	3
Storm:	
Average storm duration (d)	3
Yearly storm magnitude (Beaufort)	2
Turbidity current:	
Sua (Pa/m)	400
Sub (Pa)	0
Entrainment constant, ea (-)	.00153
Entrainment constant, eb (-)	.00204
Drag coefficient (-)	.004
Internal friction angle (deg)	20
Width of channel (m)	1000
Length of channel (km)	10
Algorithm (inflow/sakura)	Inflow
Erosion:	
Reach of highest order stream	10,000
Relief of highest order stream	25
Method (diffusion/slope)	Diffusion

References Cited Only in Appendix

Bassingthwaighte, J., and Raymond, G. 1994. Evaluating rescaled ranged analysis for time series. Ann. Biomed. Eng. 22:432–444.

Bookstein, F. L. 1988. Random-walk and the biometrics of morphological characters. Evol. Biol. 23:369–398.

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

Siegel, S., and Castellan, N. 1988. Nonparametric statistics for the behavioral sciences. 2nd ed. New York, McGraw-Hill, 399 p.

Turcotte, D. L. 1997. Fractals and chaos in geology and geophysics. 2nd ed. Cambridge, Cambridge University Press, 412 p.

Wallis, J., and Matalas, N. 1969. Small sample properties of H and K-estimators of the Hurst coefficient H. Water Resour. Res. 5:1583–1594.