Paleobiology, 26(3), 2000, pp. 319–329

MATTERS OF THE RECORD

Understanding the dynamics of trends within evolving lineages

John Alroy

National Center for Ecological Analysis and Synthesis, University of California, 735 State Street,
Santa Barbara, California 93101. E-mail: alroy@nceas.ucsb.edu

Accepted: 3 May 2000

Introduction

The study of evolutionary trends is one of the oldest and most intriguing topics in evolutionary biology and paleobiology (McNamara 1990). Workers since Cuvier, Lyell, and Owen have wanted to know if the fossil record demonstrates ''progression'' within temporal sequences of related organisms. Regardless of whether changes in the average values of morphological characters are progressive in any meaningful sense, these changes are still of great interest. In practice, questions about trends are most commonly framed by paleontologists in terms of ''complexity'' (however defined) or body size (McShea 1998a).

Research on evolutionary trends has intensified over the last few years, bringing several fundamental conceptual issues to a head. Here I analyze these conceptual problems, concluding that paleontologists should largely abandon a key method used in most studies: comparing among-species morphological distributions in successive time slices, irrespective of phylogenetic patterns. These data supposedly may distinguish random evolution, constant directional trends, and diffusion away from morphological boundaries. However, many other simple evolutionary dynamics may result in the same nonrandom trends, making it difficult to distinguish qualitatively distinct mechanisms using time-slice data. Thus, I will try to show that comparisons between ancestral and descendant morphologies need to be made instead.

The fundamental problem here is not so much mathematical or statistical as conceptual. Null hypotheses have not always been framed rigorously, and the logical connection between underlying hypotheses and methods for testing these hypotheses has sometimes been weak. Although paleontologists have long since embraced the use of null models, they still do not employ a single, straightforward definition of ''randomness'' (Eble 1999). As a result, different authors use different null hypotheses, and depending on their conceptual outlooks they may interpret the same kinds of patterns as either confirming or refuting the existence of trends. Considering the volume of literature and the century-long genealogy of the subject, this lingering panoply of viewpoints may come as a surprise.

Two points of confusion are responsible for the lack of agreement. First, it has been recognized for more than two decades now that morphological trends may result from processes operating at different levels of the phylogenetic hierarchy (Stanley 1975; Hull 1980). However, it is not always easy to say whether a particular pattern shows that nonrandom evolution is occurring at a particular hierarchical level (Hull 1980; Hoffman 1989). Second, most workers have tried to fit evolutionary patterns into binary categories of randomness and nonrandomness. However, nearly identical evolutionary trends may be generated by a surprisingly large number of qualitatively distinct mechanisms—even within a single hierarchical level.

The real question, then, is not whether evolution is random, but instead exactly what mechanisms do govern demonstrable trends, and at what hierarchical level these mechanisms operate. The ancestor-descendant method advocated here may solve both problems at once. It should often point to a proximate mechanism that generates a trend if there is

0094-8373/00/2603-0001/$1.00

one, illustrating such a mechanism clearly even if it is quite complex. Because the data only apply to within-lineage patterns, the approach also avoids the common error of conflating processes operating at different hierarchical levels.

Within- and Among-Lineage Trends

There are many ways to get from one distribution of morphologies at one time to another at a later time (Stanley 1975; McNamara 1990; McShea 1998a,b). For example, if most members of a clade are small early on, but most are large at a later date, that could be a random pattern, or it could be because either (1) natural selection (or conceivably mutation pressure, or an ecophenotypic response) has operated within evolving populations to increase mean body size; (2) species with small average body size are unlikely to give rise to daughter species; or (3) small species are more likely to go extinct.

These scenarios break down into two even narrower categories: trends (random or not) that operate within evolving populations, and trends that operate by sorting among species. The same within- vs. among-lineage distinction was recognized decades ago and was the subject of intense theoretical debate in the 1980s (Hull 1980; Hoffman 1989; McNamara 1990). Surprisingly, relatively little empirical research followed directly from this conceptual advance.

One possible explanation is the difficulty of obtaining high-quality data. To show unequivocally that a nonrandom pattern resulted mainly from a mechanism operating within, but not among, lineages, one would have to not only examine changes within lineages but quantify speciation and extinction rates across the same clade. So one would have to study a clade with many species, because otherwise quantifying speciation and extinction rates precisely would be impossible.

Very few studies have met these criteria. Norris's work on foraminiferans came close, but he did not have access to phylogenetic data that would have allowed direct analyses of evolving lineages (Norris 1991). However, Wagner's later study of early Paleozoic snails did include phylogenetic data, and it did show nonrandom evolution within lineages after controlling for apparent sorting among lineages (Wagner 1996). Likewise, recent results on body mass in mammals (Alroy 1998) and suture complexity in ammonites (Saunders et al. 1999) also can only be interpreted as showing nonrandom evolution within lineages. Still, though, these three studies didn't try to quantify the relative importance of within- and among-lineage factors—they merely sought to establish the fact that within-lineage trends are real.

"Passive" and "Driven" Trends

Over the last decade there have been many other studies of trends in morphospaces (i.e., distributions of species across quantitatively defined morphological gradients). The morphospace tradition was initially inspired by Raup (1966), and it has flowered greatly with improvements in multivariate statistical methods and computational technology. However, most morphospace studies have yielded relatively ambiguous results. I would argue that the reason is a conceptual hurdle. Instead of focusing on within- vs. among-lineage patterns, the literature on morphological trends typically invokes a very different argument presented by Stanley (1973), Fisher (1986), and McShea (1994).

As McShea (1994) explains so clearly, a nonrandom trend could result from one of two process. First, there could be a constant force (such as a bias in the probability of change) that operates across all morphologies (an ''active'' or ''driven'' trend). Second, there could be a force that is variable across the range of morphologies (such as an absorbing or reflecting boundary), which might truncate the distribution of morphologies (''passive diffusion''). For example, there may be an absolute minimum body size that pertains to all members of a clade. Some workers have indeed recognized that such a boundary could be maintained by either within- or among-lineage factors (e.g., McShea 1998b), but for the most part they have not stressed the point. Even more importantly, many workers have interpreted McShea's 1994 paper to mean that passive diffusion is no more interesting than purely random evolution, leaving only driven trends as evidence of genuinely interesting evolutionary processes.

McShea (1994) proposed a ''subclade'' test that attempted to distinguish passive and driven trends by contrasting static distributions of morphologies observed at different times. One strength of this test is that it does not require a detailed phylogeny, unlike the ancestor-descendant comparisons that McShea recognized as the alternative basis of such analyses. Although most obviously applicable to paleontological data sets, McShea's test has drawn so much attention that it has even been employed by comparative biologists using neontological data (e.g., Maurer 1998).

Many studies have operated within the passive/driven trend paradigm, but few of them have found conclusive results. Passive diffusion without differential turnover seemed to operate in a study of ammonite suture complexity (Boyajian and Lutz 1992), but later, more comprehensive studies of shell geometry found strong hints of more complicated nonrandom patterns and disagreed on whether there are strongly directional trends (Dommergues et al. 1996) or diffusive trends away from boundaries (Saunders and Work 1996; but see Saunders et al. 1999). An elegant discussion of brachiopod biomechanics suggested both kinds of trends, but didn't document these claims by examining individual evolutionary lineages (Carlson 1992). A comprehensive analysis of body size in Late Cretaceous bivalves seemed to show no trends at all (Jablonski 1996). Work on foraminiferans showed sorting among lineages, but interpreted trends within lineages as resulting from sampling biases, or at best suggesting passive diffusion (Arnold et al. 1995). A study of trilobite body size suggested differential turnover, but didn't try to document trends within lineages (Trammer and Kaim 1997). None of these studies employed the methodology I am about to discuss; nor do quite a few others that I have no space to detail.

Ancestor-Descendant Comparisons: A Window on Evolutionary Dynamics

Does using McShea's distinction of random, passive, and driven mechanisms (McShea 1994) solve the riddle of classifying evolutionary trends? As recognized by McShea (1998b), it seems instead that the menagerie of mechanisms is far too diverse to be broken down into just three categories. Instead, the best way to illustrate this point may be to plot graphs that compare the morphology of an ancestor (the starting point of a trend) with the change in morphology from the ancestor to the descendant (the direction of the trend [Alroy 1998]). Details of this ''branch breaking'' method are given in Figure 1.

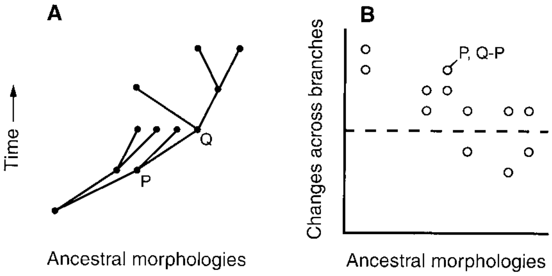

FIGURE 1. A simple case in which a trend within lineages plays out across a phylogenetic tree. The trend constitutes a bias in the direction of change across all branches of a tree. The ''small'' or ''simple'' morphologies at the left-hand side of a morphological spectrum tend to move quickly toward more ''large'' or ''complex'' morphologies to the right (A); but once those morphologies are attained, there is little additional change. The fact that the same pattern is seen across independent lineages shows that the trend is not random, but instead the deterministic result of underlying dynamic rules. The rules are best illustrated by ''breaking'' the branches to produce a plot (B) of differences between descendants and ancestors (changes across each branch) against ancestral morphologies (morphologies at the start of each branch). For example, the morphology at node P is 5 and that at its descendant node Q is 8, so the branch is represented in the change-vs.-ancestor plot by the point labeled ''P, Q-P'' at the coordinate 5,3. The plot suggests a point equilibrium, as in Figure 2C, but it may be consistent with more complex dynamics (e.g., Fig. 2E,I,K). A more traditional analysis would plot frequencies of different morphologies seen during different time periods, which would fail to show clearly whether the pattern was due to random diffusion, sorting of different lineages, or (as in this case) deterministic trends within lineages.

Before continuing, the key features of this kind of a plot need to be justified. First, one might ask why a simple regression analysis of ancestor and descendant morphologies wouldn't suffice instead. That would avoid having to worry about quantifying changes in traits through time. The reason not to do this is that quantitative characters like lengths, areas, and volumes often evolve quite slowly, so closely related species will have very similar values. Thus, regression plots would show such strong linear relationships that the subtle departures from linearity I am about to discuss would be obscured. Instead, the best way to highlight these departures is to difference one variable, i.e., replace the descendant values with differences between descendant and ancestral values.

Second, one might ask why only the descendant values are to be differenced. After all, comparative biology is nowadays much focused on plots that contrast changes along pairs of collateral branches stemming from a common ancestor (sister-species comparisons [e.g., Miles and Dunham 1993; Butler and Losos 1997]). There are three reasons not to employ that methodology here: (1) Paleobiologists often really do deal with ancestor-descendant relationships and not just sister-species relationships. The branch-breaking method would work even if all species were linked in one long anagenetic chain with no cladogenesis at all. By contrast, sister-species comparisons throw out the information in anagenetic sequences. (2) Comparative analyses typically focus on pairs of characters, not just a single character, and the standard methods are not designed to handle the latter problem. (3) More specifically, sister-species comparisons can obliterate all traces of complex dynamics by removing any direct information about an ancestor's morphology. For example, in an analysis of body mass one would lump together all changes from shrew-sized ancestors with all changes from elephant-sized ancestors, asking only if changes are strongly correlated in collateral branches and not if the average amount of change varied with body mass in any systematic way. Thus, sister-species examinations can tell us only about linear correlations and nothing about the many imaginable, qualitatively distinct dynamic rules that may be in operation.

Third, there is the more practical question of where the values are to be obtained. In comparative biology, the ancestral values might be estimated from the observed distribution of character states across a phylogeny by using (for example) least-squares parsimony (Felsenstein *in* Huey and Bennett 1987; Maddison 1991). In paleobiology, one could either use the same sort of phylogenetic estimates or work with ancestor-descendant pairs of named morphospecies, as identified by methods like stratophenetics or stratocladistics (Smith 1994; Wagner 1998). The exact source of the data is a secondary methodological issue not treated in this paper.

Finally, there is the related question of just what is meant by taking differences between descendants and ancestors. Are these differences evolutionary rates per se? Again, the point of this paper is not to dwell on how exactly one should quantify such rates (Gingerich 1983, 1993): the graphs could represent darwins (changes per unit time), haldanes (changes per unit generation), or simple differences, which are just rates per speciation event. However, I will note that if lineages are largely static in between speciation events, then the latter, simple-most method is probably to be preferred. Computing darwins or haldanes requires dividing through by the amount of time elapsed, which is problematic because the temporal precision of paleontological timescales is often as long as, or longer than, the median duration of individual species. Therefore, this kind of division can introduce random error into a comparison and thereby obscure a real evolutionary signal. For reasons such as these I used simple differences in an earlier empirical study (Alroy 1998).

A Menagerie of Evolutionary Trends

Using a change-vs.-ancestral state plot, one may envision at least twelve simple, but qualitatively different, outcomes (Fig. 2). In the most simple hypothetical cases, there is no relationship between ancestral values and changes (Fig. 2A,B). If the average change is zero (Fig. 2A), evolution truly appears to be random. By contrast, any non-zero average value results in a nonrandom trend that operates across all morphologies (Fig. 2B)—so as McShea (1998b) put it, the morphospace (or more generally ''state space'') is ''unstructured.'' These two plots may show what McShea (1994) meant by purely random and driven trends.

The next two graphs illustrate a linear relationship between ancestral values and

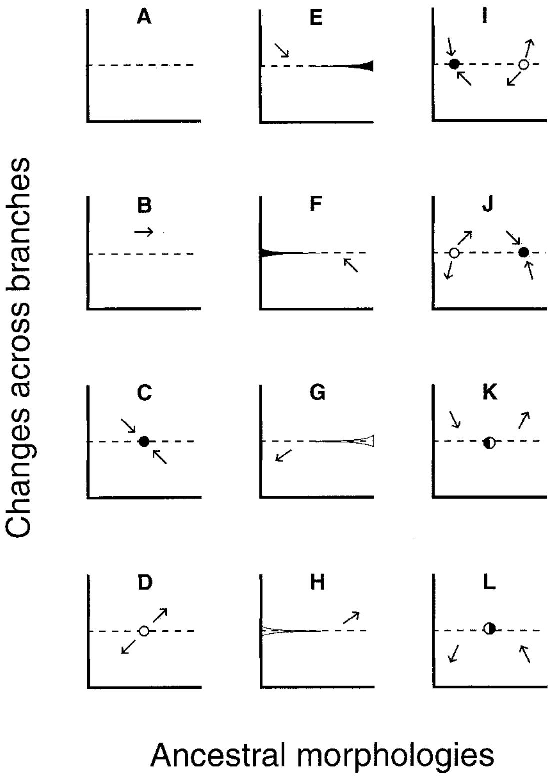

FIGURE 2. Sketches of qualitatively distinct dynamics of evolutionary change. Ancestral morphological values (e.g., size or complexity) are contrasted with differences between descendants and ancestors (e.g., amounts of evolutionary change in size or complexity). Dashed lines indicate evolutionary changes of zero; filled circles and tapering shapes are stable equilibria; open circles and tapering shapes are unstable equilibria; arrows show expected directions of change toward stable equilibria or away from unstable equilibria; gray shapes suggest possible scatters of data points. A, Random change. B, Constant, directed bias. C, Single stable point equilibrium. D, Single unstable point equilibrium. E, Stable equilibrium zone implying a "lower bound." F, Stable equilibrium zone implying an "upper bound." G, Unstable equilibrium zone implying a "lower cliff." H, Unstable equilibrium zone implying an "upper cliff." I, Double point equilibria: lower is stable, upper is unstable. J, Double point equilibria: lower is unstable, upper is stable. K, Point equilibrium that is stable for low values but unstable for high values (i.e., a saddle). L, Point equilibrium (saddle) that is unstable for low values but stable for high values.

changes. In other words, these are "structured" state spaces (McShea 1998b). In Figure 2C, very low values are followed by large positive changes, and high values by negative changes. As a result, all morphologies will converge on a single value. This point in the graph can be thought of as a "stable equilibrium," "optimum," or "point attractor." The implied dynamic is closely analogous to the carrying capacity of a population experiencing logistic growth (in ecology) (May 1974), the optimal allelic frequency when there is heterozygous advantage or frequency-dependent selection (in genetics) (Lewontin and Kojima 1960), or an Ornstein-Uhlenbeck attractor (in physics) (Martins 1994). Searching for stable equilibria is called a "stability analysis" in much of the literature on dynamical systems. The change-vs.-ancestor comparison method is just one approach to this problem of many, but it is uniquely appropriate because other methods usually deal with linear time series of a single state variable, not the complex branching patterns that generate evolutionary trends.

The following plot (Fig. 2D) seems similar, and it does include an equilibrium point—the x-intercept of the line of expected change, i.e., where the expected change from ancestor to descendant is zero. However, this is an unstable equilibrium or "repeller": small values lead to negative expected changes, so they result in runaway trends away from the equilibrium, and likewise for large values. Qualitatively, these two dynamics are very different from each other. The first dynamic implies that most evolving lineages should converge on a single value, whereas the second implies that there should be rapidly accelerating trends leading away from the same point toward either higher or lower values.

McShea's passive/driven formality does not directly address the possible existence of point equilibria, stable or not. What, then, does he mean by "passive" trends? The next four cartoons show cases in which expected changes are more or less an exponential function of ancestral values. Two of these dynamics result in stable solutions (Fig. 2E,F): extreme morphologies will tend to move quickly toward high values (in the first case) or low values (in the second). However, the stable solutions don't correspond to a single point; instead, there is an "equilibrium zone" (Alroy 1998) within which small evolutionary changes are effectively random. Both of these dynamics can be thought of as "bounded" morphospaces that are governed by "passive" diffusion. Specifically, a "lower boundary" can be thought of as the point on the left-hand side of Figure 2E where the expected change is so great that in practice the scatter around this change includes only positive values. An analogous upper boundary is created by the dynamic in Figure 2F.

The next six types of dynamics don't fit into the random/passive/driven typology in such an obvious way. Figures 2G and H show unstable equilibrium zones; evolution is random for most morphologies, but extreme morphologies tend to run away from these zones at ever increasing rates (i.e., they ''fall off a cliff'' instead of ''hitting a wall''). The last row of graphs shows more complex dynamics in which there may be both stable and unstable point equilibria (Fig. 2I,J) or else equilibrium zones that are stable on one side and unstable on the other (i.e., saddle points; Fig. 2K,L). All six patterns involve unstable equilibria, which are the very opposite of optima: lineages tend to diverge from an equilibrium at ever increasing rates. Such dynamics may seem so exotic as to be biologically impossible, but in fact there is a well-established body of literature concerning real-world cases of ''runaway'' evolution—which are just trends away from unstable equilibria (Endler and Basolo 1998).

All twelve dynamics can be inferred from real data, for example, by using simple polynomial regression methods that never involve more than three explanatory parameters (Alroy 1998). Random evolution (Fig. 2A) requires no parameters; a simple driven trend (Fig. 2B) requires one, the mean change; single point attractors and repellers (Fig. 2C,D) require two, the slope and intercept of the regression line; exponential relationships (Fig. 2E–H) also require two, the rate of decay and a point to fix the curve along the x-axis; and the more complex dynamics (Fig. 2I–L) all can be derived using a three-parameter quadratic regression equation, or using spline fits of similar complexity.

The fact that each qualitative pattern corresponds to a process with a precisely defined minimum number of explanatory parameters suggests that we can classify within-lineage trends into four simple categories, or ''rungs,'' with each rung corresponding to a set of dynamics requiring the same number of descriptive parameters. So rung 0 is a random walk (Fig. 2A), rung 1 is an unstructured bias operating across the state space (Fig. 2B), and so on.

Note that in computing these counts of parameters, I have simply omitted the extra term that would summarize the error in the model. As an example, in the case of random evolution at least one parameter would be needed to describe the change variable's variance around its mean value of zero. So this distribution does require descriptive parameters in a trivial sense, but none that have anything to do with explaining the relationship between the changes and ancestral states. Because the whole point of the exercise is to see how far a given dynamic departs from randomness, excluding the error term only makes sense.

Surprisingly, relatively few studies have used within-lineage ancestor-descendant plots like the one in Figure 2 to analyze real data. One of them (Alroy 1998) found a dynamic that didn't correspond to any of the simple cases first highlighted by McShea (1994), or even to the simple point-attractor dynamic—instead, it seemed to show the peculiar double-equilibrium dynamic of Figure 2J, with its implication of a near-runaway trend.

Reconciling Terminology: Is Any Trend Really "Passive"?

It seems clear that in the terminology of McShea (1994), Figure 2A is ''random,'' Figure 2B is ''driven,'' and Figure 2E and F show ''passive diffusion.'' Arguably, the more exotic dynamics of Figure 2G–L also all fall into the ''passive'' category. All of these exotica resemble the examples of ''structured state spaces'' illustrated by McShea (1998b), in which ''small-scale dynamics'' like turnover rates and within-lineage trends follow markedly different rules in different parts of a morphological gradient. McShea (1998b) largely abandoned the ''passive/driven'' distinction in analyzing these sorts of complex dynamics, but the fact that most of them involve stable attractors that create apparent boundaries (e.g., Fig. 2I–L) does suggest that classifying them as cases of passive diffusion would be consistent with his viewpoint.

Even if the classification of rung 2 and 3 dynamics as ''passive diffusion'' is fair to McShea (1994, 1998b), it is dangerous for at least three reasons. First, one might argue that purely random evolution is also passive because the random change in a distribution that inevitably results—and no unconstrained distribution can evolve without changing—is surely not driven. So on this account rungs 0, 2, and 3 are all in the same ''passive'' category. This mapping would be hopelessly confusing because the opposed ''driven'' category would now apply only to mechanisms that occupy an intermediate position (rung 1) in the overall gradient of dynamic complexity.

Second, many other workers assume that even if there are such things as morphological boundaries (i.e., state-space structures), these ''passive'' rung 2 and 3 trends are still somehow random (but see McShea 1998b). Of course, this really makes no sense if within-lineage factors are important: all of the mechanisms that might create such trends do involve nonrandom, nonlinear dynamics, with attractors that are even more complex than simple point attractors. One could also construct an analogous argument that if ''passive'' diffusion results from a boundary that is implemented by among-lineage biases in turnover rates (McShea 1998b), this is anything but a simple random process. So in fact, all of the complex dynamics do involve locally strong and active ''driving'' of lineages across the morphospace—even though the strength and direction of driving varies depending on the starting point. Essentially, calling rung 1 trends driven and rung 2 and 3 trends passive makes it sound as though the former are more interesting, when the opposite is really true.

Finally, no matter how one wants to draw analogies, McShea's random/passive/driven alternatives apply cleanly to only four of the twelve dynamics (Fig. 2A,B,E,F). Any mapping of the other eight dynamics into his categories is problematic. The three-way split also fails to emphasize one solution that may be very widespread: the point attractor (Fig. 2C), which should exist (for example) whenever there is a single biomechanical solution to a functional problem. McShea (1998b: Fig. 6.4B) illustrated a case that essentially does involve such a dynamic, but in his terminology this is just one of many ''somewhat structured'' morphospaces. The four-rung classification does emphasize the distinct status of simple point attractors, and it makes their intermediate relative degree of complexity quite clear.

How, then, can one reconcile the random/passive/driven distinction with the complexity gradient advocated here? I think the task is impossible. The above-mentioned objections to leaving McShea's terminology intact are simply too strong. There is an alternative: one could modify his definitions by, say, equating rung 0 with passive diffusion, rung 1 with simple driven trends, and rungs 2 and 3 with driven trends operating in a structured state space. However, this kind of retroactive semantic juggling would generate even more confusion. Thus, despite great misgivings I advocate abandoning McShea's terminology.

What Patterns Result from Different Evolutionary Dynamics?

Plots like those in Figure 2 are almost never encountered in the paleontological or comparative biological literature. Instead, most paleontological studies plot morphological distributions directly against time, or present a series of histograms or scatter plots depicting the same morphospace in different time slices (for example, all Cretaceous species in one plot and all Paleocene species in another). Many recent papers have employed this approach, including studies of palynomorphs (Lupia 1999), foraminiferans (Norris 1991; Arnold et al. 1995), trilobites (Foote 1991; Sundberg 1996; Smith and Lieberman 1999), crustaceans (Wills 1998), bivalves (Jablonski 1996), rostroconchs (Wagner 1997), gastropods (Roy 1994, 1996; Wagner 1996), ammonites (Boyajian and Lutz 1992; Dommergues et al. 1996; Saunders and Work 1996; Saunders et al. 1999), brachiopods (Carlson 1992), blastozoans (Foote 1992), crinoids (Foote 1995), echinoids (Eble 1998), ungulates (Jernvall et al. 1996), and even theropod dinosaurs (Gatesy and Middleton 1997). The distributions (often derived from multivariate analysis of complex

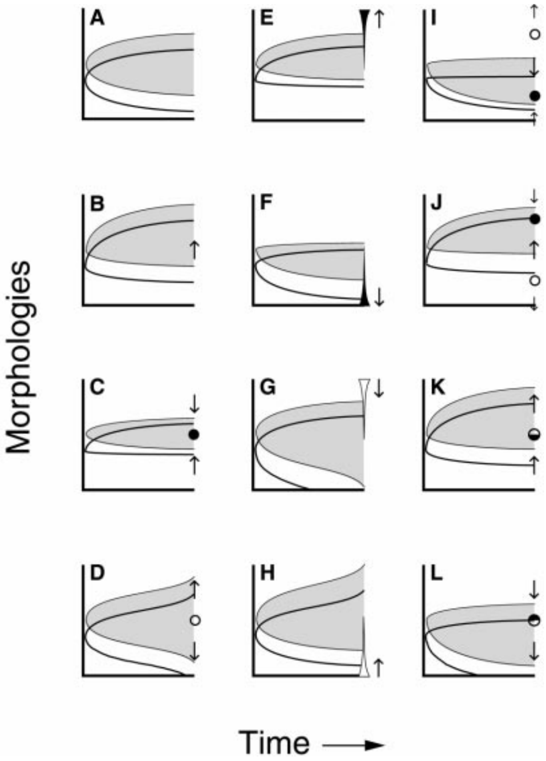

FIGURE 3. Sketched patterns of progressive morphospace occupation implied by distinct evolutionary dynamics. Each panel corresponds to an underlying dynamic depicted in Figure 2. Morphology is contrasted with time, which moves from left to right. Evolution is assumed to have a large stochastic component, but deterministic dynamical rules are partially responsible for governing the trends. Open envelopes and lightly shaded envelopes suggest the ranges of morphologies seen within two different hypothetical clades that are differentiated only by their starting points. Despite evolving under the same dynamical rules for the same amount of time, the final distributions of the two clades almost never converge completely when starting points differ; and despite each panel's showing a distinct dynamic, many patterns are very similar.

morphologies) may be reduced to means, variances, minima, maxima, or ranges.

Most authors hope to infer evolutionary dynamics from these raw patterns. Optimally, there might be a one-to-one correspondence between dynamic mechanisms and temporal patterns: driven trends might generate one pattern, point attractors a second, and so on. But in fact, if evolution has a large stochastic component, then many of the qualitatively very different dynamics will map to similar or identical temporal distributions (Fig. 3). Random trends, simple driven trends, and trends driven by single, unstable equilibrium points all will generate steady increases in variance over long stretches of time (Fig. 3A,B,D,G,H). Apparently directed trends (Fig. 3B) could result also from distributions that converge slowly on a stable equilibrium from a different starting value (Fig. 3C,E,F). But if the starting value happens to be close to the equilibrium, no driven trend will appear. Likewise, in case after case the same dynamic rules generate qualitatively different temporal trends depending on the starting point; just look at the ''lower cliff'' model, which could result in either a very strong trend or something resembling pure randomness (Fig. 3G).

Of course, all of these inferences are based on hand-drawn plots; simulation studies might show that certain kinds of dynamics might in fact be distinguishable using temporal distribution data. However, it seems a safe bet that as helpful as they might be, temporal distributions still may obscure processes that can be perceived more directly by examining a change-vs.-ancestor plot—if only because the latter would restrict the analysis to within-lineage evolutionary patterns, instead of conflating the among- and within-lineage outcomes that sum up to create a broad-brush temporal sequence of distributions.

Instead of natively investigating morphology-by-time plots, many recent authors employ a more sophisticated argument, following Foote (1991, 1992, 1995). If the range of morphology expands much more rapidly than the rate of taxonomic diversification per se, then one might infer the existence of driven trends, perhaps toward an equilibrium. However, this style of analysis is handicapped by its inability to distinguish within- from among-lineage dynamics, or even (assuming optimistically that the latter is not important) different kinds of within-lineage ''driven'' trends. After all, many qualitatively different dynamics can generate a rapid initial burst of morphological change (e.g., open envelope in Fig. 3C,F,J; lightly shaded envelope in Fig. 3E,I). So the taxonomic diversity vs. morphological disparity approach may help to rule out some possibilities, but it cannot provide positive evidence in favor of others.

Where to Now?

The implication seems strong: plots of static distributions in different time slices (Fig. 3) don't really tell us very much about underlying evolutionary dynamics (Fig. 2). These temporal-trend plots (Fig. 3) often illustrate confused signals of within- and among-lineage trends, have no necessary implications for microevolution, and can't even in theory distinguish such fundamentally different dynamical systems as, say, point attractors (Figs. 2C, 3C), unstable equilibrium zones (Figs. 2H, 3H), and saddles (Figs. 2K, 3K). By contrast, dynamic plots (Fig. 2) pertain only to within-lineage trends, have direct implications for microevolutionary mechanisms, and can be distinguished using real data even when the dynamics are complex and nonlinear—as long as the sample sizes are large enough to overcome stochastic effects (Alroy 1998). As shown above, the counterargument that time-slice data may at least reveal ''boundaries'' in morphospaces is both conceptually misleading and irrelevant: ''boundaries'' are nothing more than certain kinds of underlying dynamics, and these dynamics are better illustrated by dynamic plots in the first place.

These conclusions followed from one empirical study (Alroy 1998) and a simplistic thought experiment that only dealt with very rudimentary dynamical models operating strictly within lineages. Imagine how much more complicated things could get. Instead of simple bell-shaped curves in these change vs. ancestor plots (Fig. 2I–L), there might be curves with multiple inflection points, which could be illustrated with higher-order polynomial fits and might imply as many equilibrium points as there are polynomial terms. There also could be circular relationships, which would imply limit cycles; or even more oddly shaped nonlinear patterns that might imply some combination of cycles and equilibria. Many well-established statistical techniques can be used to study these kinds of complex dynamical systems, but most of these methods focus on analyzing linear, temporal sequences of data points (e.g., Sugihara and May 1990). Because no natural, linear sequence of all the changes along the phylogenetic branches in a data set can be constructed, applying such methods may be problematic.

On top of worrying about complex, nonlinear, within-lineage dynamics, we might also try to model among-lineage mechanisms—differential speciation and extinction rates in different parts of the morphospace. That would create a whole different set of methodological problems. At the very least, as mentioned earlier one would have to assemble a very large data set to exactly quantify the relevant turnover rates. Moreover, extinction rates are particularly hard to quantify because they are a function of sampling: as more fossils are collected, age ranges of known species lengthen and previously unsampled species are found. There are good ways to handle this problem (Marshall 1990, 1997; Foote and Raup 1996; Foote 1997), but the methods have their own biases. Also, analyses of comparative data for Recent groups without good fossil records would simply have to put the extinction problem aside. The situation is even more complicated for speciation rates: to find the rates for a specific subregion of a morphospace, one should at least have a rough approximation of the ancestral morphologies, which means having some basic phylogenetic data.

To summarize, dynamic analyses relating state variables (ancestral morphologies) to changes in state variables (trends within lineages) can uncover a rich smorgasbord of evolutionary mechanisms. Despite this, a large body of extremely sophisticated ''empirical morphospace'' literature has placed a strong emphasis on morphology-by-time plots instead of morphologic change-by-ancestral-morphology plots. Most of these studies unfortunately do not have access to the kind of detailed, species-level phylogenetic data one ultimately would need to untangle within- from among-lineage dynamics, and to discriminate the wild variety of within-lineage dynamics illustrated in Figures 2 and 3. However, a small but rapidly growing cohort of studies like those of Wagner (1996), Saunders et al. (1999), and Smith and Lieberman (1999) does include the requisite data, suggesting that there are good reasons to be optimistic.

Caveats

With phylogenetic and morphometric studies becoming more and more common both in paleontology and in comparative biology, it now seems reasonable to call for a wedding of these studies to produce dynamic analyses of within- and among-lineage trends. However, this is not to say that all of the important conceptual and statistical problems have been resolved.

First, there is a legitimate concern that the kinds of analyses discussed here do not really get at the ultimate—as opposed to proximate—causes of evolutionary trends. Just because there is a stable point attractor in an evolutionary system, that does not mean we know why the attractor exists, or even if it specifically implies, say, some form of biomechanical optimality. Instead, there could be developmental constraints that make it difficult or impossible to create extreme morphologies; there could be morphological limits imposed by biological interactions, such as competition, predation, or parasitism; or there could be particular abiotic environmental factors that make a morphology optimal only during the time and place that is being studied—perhaps a different optimum would exist, say, in a different climate regime.

Second, there are serious problems with the data and statistical methods that are needed to study evolutionary dynamics. Highly resolved phylogenies are rare, and phylogenetic methods are just as hotly debated in paleontology as elsewhere (e.g., Smith 1994; Wagner 1998). Morphometric data sets come with built-in sampling errors and are sometimes an arbitrary function of the variables that have been chosen. Taking differences between ancestral and descendant values could compound any error that may exist, a problem that intensifies as the quality of a phylogeny decreases.

Finally, the method of plotting changes within lineages against ancestral values has its own unique bias: the noise introduced by poor phylogenies and poor measurements may specifically result in negative linear patterns that are spurious. Incorrectly low ancestral values will on average be followed by positive changes back to the mean, and vice versa, so with much noise the mean will automatically appear to be an ''attractor.'' This ''regression to the mean'' problem (Alroy 1998) has not really been solved, either for evolutionary study systems or for related ones such as those concerning the dynamics of ecological populations (Dennis and Taper 1994). The one attempt at imposing a correction factor that would account for evolutionary regressions to the mean rested on a briefly sketched equation that has not yet been documented fully (Alroy 1998).

Despite all of this, paleobiology and comparative biology are now at an exciting crossroads. Large, high-quality data sets for many different groups will soon become common place, and the basic conceptual problems really are not so severe. We may yet choose the right path toward a greater understanding of evolutionary trends.

Acknowledgments

I thank P. Amarasekare, G. Eble, J. Hunter, J. Losos, B. Maurer, D. McShea, H. Sims, and P. Wagner for comments. Research was supported by the National Center for Ecological Analysis and Synthesis, a center funded by the National Science Foundation (grant DEB-94-21535); the University of California, Santa Barbara; the California Resources Agency; and the California Environmental Protection Agency.

Literature Cited

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

Arnold, A. J., D. C. Kelly, and W. C. Parker. 1995. Causality and Cope's rule—evidence from the planktonic Foraminifera. Journal of Paleontology 69:203–210.

Boyajian, G., and T. Lutz. 1992. Evolution of biological complexity and its relation to taxonomic longevity in the Ammonoidea. Geology 20:983–986.

Butler, M. A., and J. B. Losos. 1997. Testing for unequal amount of evolution in a continuous character on different branches of a phylogenetic tree using linear and squared-change parsimony: an example using Lesser Antillean *Anolis* lizards. Evolution 51:1623–1635.

Carlson, S. J. 1992. Evolutionary trends in the articulate brachiopod hinge mechanism. Paleobiology 18:344–366.

Dennis, B., and M. L. Taper. 1994. Density dependence in time series observations of natural populations—estimation and testing. Ecological Monographs 64:205–224.

Dommergues, J.-L., B. Laurin, and C. Meister. 1996. Evolution of ammonoid morphospace during the Early Jurassic radiation. Paleobiology 22:219–240.

Eble, G. J. 1998. Diversification of disasteroids, holasteroids and spatangoids in the Mesozoic. Pp. 629–638 *in* R. Mooi, and M. Telford, eds. Echinoderms. Balkema, Rotterdam.

———. 1999. On the dual nature of chance in evolutionary biology and paleobiology. Paleobiology 25:75–87.

Endler, J. A., and A. L. Basolo. 1998. Sensory ecology, receiver biases and sexual selection. Trends in Ecology and Evolution 13:415–420.

Fisher, D. C. 1986. Progress in organismal design. Pp. 99–117 *in* D. M. Raup, and D. Jablonski, eds. Patterns and processes in the history of life. Springer, Berlin.

Foote, M. 1991. Morphological patterns of diversification—examples from trilobites. Palaeontology 34:461–485.

———. 1992. Paleozoic record of morphological diversification in blastozoan echinoderms. Proceedings of the National Academy of Sciences USA 89:7325–7329.

———. 1995. Morphological diversification of Paleozoic crinoids. Paleobiology 21:273–299.

———. 1997. Estimating taxonomic durations and preservation probability. Paleobiology 23:278–300.

Foote, M., and D. M. Raup. 1996. Fossil preservation and the stratigraphic ranges of taxa. Paleobiology 22:121–140.

Gatesy, S. M., and K. M. Middleton. 1997. Bipedalism, flight, and the evolution of theropod locomotor diversity. Journal of Vertebrate Paleontology 17:308–329.

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

———. 1993. Quantification and comparison of evolutionary rates. American Journal of Science 293A:453–478.

Hoffman, A. 1989. Arguments on evolution: a paleontologist's perspective. Oxford University Press, New York.

Huey, R. B., and A. F. Bennett. 1987. Phylogenetic studies of coadaptation—preferred temperatures versus optimal performance temperatures of lizards. Evolution 41:1098–1115.

Hull, D. L. 1980. Individuality and selection. Annual Review of Ecology and Systematics 11:311–332.

Jablonski, D. 1996. Body size and macroevolution. Pp. 256–289 *in* D. Jablonski, D. H. Erwin, and J. H. Lipps, eds. Evolutionary paleobiology. University of Chicago Press, Chicago.

Jernvall, J., J. P. Hunter, and M. Fortelius. 1996. Molar tooth diversity, disparity, and ecology in Cenozoic ungulates. Science 274:1489–1492.

Lewontin, R. C., and K. Kojima. 1960. The evolutionary dynamics of complex polymorphisms. Evolution 14:458–472.

Lupia, R. 1999. Discordant morphological disparity and taxonomic diversity during the Cretaceous angiosperm radiation: North American pollen record. Paleobiology 25:1–28.

Maddison, W. P. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. Systematic Zoology 40:304–314.

Marshall, C. R. 1990. Confidence intervals on stratigraphic ranges. Paleobiology 16:1–10.

———. 1997. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. Paleobiology 23: 165–173.

Martins, E. P. 1994. Estimating the rate of phenotypic evolution from comparative data. American Naturalist 144:193–209.

Maurer, B. A. 1998. The evolution of body size in birds. I. Evidence for non-random diversification. Evolutionary Ecology 12:925–934.

May, R. M. 1974. Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos. Science 186:645–647.

McNamara, K. J., ed. 1990. Evolutionary trends. University of Arizona Press, Tucson.

McShea, D. W. 1994. Mechanisms of large-scale evolutionary trends. Evolution 48:1747–1763.

———. 1998a. Possible largest-scale trends in organismal evolution: eight ''live hypotheses. '' Annual Review of Ecology and Systematics 29:293–318.

———. 1998b. Dynamics of diversification in state space. Pp. 91–108 *in* M. L. McKinney, and J. A. Drake, eds. Biodiversity dynamics: turnover of populations, taxa, and communities. Columbia University Press, New York.

Miles, D. B., and A. E. Dunham. 1993. Historical perspectives in ecology and evolutionary biology—the use of phylogenetic comparative analyses. Annual Review of Ecology and Systematics 24:587–619.

Norris, R. D. 1991. Biased extinction and evolutionary trends. Paleobiology 17:388–399.

Raup, D. M. 1966. Geometric analysis of shell coiling: general problems. Journal of Paleontology 40:1178–1190.

Roy, K. 1994. Effects of the Mesozoic marine revolution on the taxonomic, morphological, and biogeographic evolution of a group—aporrhaid gastropods during the Mesozoic. Paleobiology 20:274–296.

———. 1996. The roles of mass extinction and biotic interaction in large-scale replacements: a reexamination using the fossil record of stromboidean gastropods. Paleobiology 22:436–452.

Saunders, W. B., and D. M. Work. 1996. Shell morphology and suture complexity in Upper Carboniferous ammonoids. Paleobiology 22:189–218.

Saunders, W. B., D. M. Work, and S. V. Nikolaeva. 1999. Evolution of complexity in Paleozoic ammonite sutures. Science 286:760–763.

Smith, A. B. 1994. Systematics and the fossil record: documenting evolutionary patterns. Blackwell Science, Oxford.

Smith, L. H., and B. S. Lieberman. 1999. Disparity and constraint in olenelloid trilobites and the Cambrian Radiation. Paleobiology 25:459–470.

Stanley, S. M. 1973. An explanation for Cope's rule. Evolution 27: 1–26.

———. 1975. A theory of evolution above the species level. Proceedings of the National Academy of Sciences USA 72:646–650.

Sugihara, G., and R. M. May. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344:734–741.

Sundberg, F. A. 1996. Morphological diversification of Ptychopariida (Trilobita) from the Marjumiid biomere (Middle and Upper Cambrian). Paleobiology 22:49–65.

Trammer, J., and A. Kaim. 1997. Body size and diversity exemplified by three trilobite clades. Acta Palaeontologica Polonica 42:1–12.

Wagner, P. J. 1996. Contrasting the underlying patterns of active trends in morphologic evolution. Evolution 50:990–1007.

———. 1997. Patterns of morphologic diversification among the Rostroconchia. Paleobiology 23:115–150.

———. 1998. A likelihood approach for evaluating estimates of phylogenetic relationships among fossil taxa. Paleobiology 24:430–449.

Wills, M. A. 1998. Crustacean disparity through the Phanerozoic: comparing morphologic and stratigraphic data. Biological Journal of the Linnean Society 65:455–500.