

## Species selection and driven mechanisms jointly generate a large-scale morphological trend in monobathrid crinoids

Carl Simpson

*Abstract.*—All evolution attributable to natural selection, at any level, is due to a causal covariance between fitness and phenotype. Over macroevolutionary time scales, species selection is one of many possible mechanisms for generating large-scale morphological trends. For species selection to sort morphology, a correlation between morphology and taxonomic diversification rate must be present. Other trend mechanisms (driven mechanisms, e.g., a bias in the direction of speciation) produce a systematic change in the mean phenotype over time. All mechanisms can co-occur. Here I demonstrate (1) an inverse correlation between diversification rate and calyx complexity that demonstrates the effect of species selection on morphology. Genera with simple calyces tend to increase in diversity, whereas genera with complex calyces have a net decrease in diversity; and (2) the presence of a driven trend mechanism in monobathrid crinoids where descendant genera tend to be simpler than their ancestors. The separate effects of these two classes of trend mechanisms can be combined by using the Price's Theorem, which partitions the contribution to the overall change in calyx complexity over time accurately among selection and driven mechanisms. Price's Theorem provides significant conceptual and methodological clarification of the contribution of multiple and interacting hierarchical mechanisms in generating large-scale trends.

*Carl Simpson. Museum für Naturkunde, Leibniz Institute for Research on Evolution and Biodiversity at the Humboldt University Berlin, Invalidenstrasse 43, D-10115 Berlin, Germany. E-mail: carl.simpson@mfn-berlin.de*

Accepted: 21 January 2010

### Introduction

Evolution by natural selection is easy. Only heritable variation in fitness is needed for entities to evolve by natural selection (Lewontin 1970). In principle, many hierarchical levels can satisfy these criteria, from selfish genetic elements up through populations of organisms to the species level and above. The key effect of natural selection is that it provides directional change in evolution. Large-scale evolutionary trends in the fossil record are persistent directional patterns with the potential to be produced by selection at the level of species or genera. Species and genera do, in fact, have a high degree of similarity between ancestor and descendants (Jablonski 1987; Hunt et al. 2005). In addition, for certain groups, fitness and aspects of species- and genus-level phenotypes are known to covary (Van Valen 1975; Jablonski 1986a,b; Liow 2006; Simpson and Harnik 2009).

We know that macroevolutionary dynamics vary and that mechanisms for high levels of selection exist. But what is unknown is how

causes of selection at different levels are coupled to morphology and how to tease apart the separate effects of selection at multiple levels. How much can species selection influence organismal-level morphological evolution when a lower level of selection is acting simultaneously? Evaluation of the role of species selection in morphological trends has been complicated by a methodological framework that focuses on only a single trend-generating process (e.g., Gould and Eldredge 1977) when in fact many can co-occur. Acceptance of species selection could only come by demonstrating either its direct opposition to lower levels of selection or the nonexistence of selection at a lower level. Both of these could be exceedingly rare, making the rejection of the null hypothesis of oppositional or zero selection nearly impossible.

In this paper I present a new framework for studying trends that includes the ability to estimate the joint contribution of species selection and driven mechanisms (including a lower-level of selection) to the overall trend,

if any. In addition, I empirically evaluate the role of species selection and driven mechanisms in the putative simplification trend in the monobathrid crinoid calyx.

I use Price's Theorem (Price 1972; Frank 1997; Rice 2004; Okasha 2007) to partition the effects of species selection and driven mechanisms into separate parameters, which I independently estimate from fossil data. Price's Theorem accurately models all selection processes, including macroevolutionary ones, and allows us to use a framework of estimation, rather than null hypothesis testing, in the study of species selection. The magnitude and direction of species selection is estimated from an observed covariance between calyx complexity and diversification rates. I estimate the magnitude and direction of driven mechanisms, which include microevolutionary change, indirectly using a new extension of McShea's (1994) subclade test. The generalized subclade test can estimate the vector of driven mechanisms by comparing the direction of skewness for subclades sampled in a time-slice and the cumulative morphological distribution of a subclade up to the time-slice.

### Data

Monobathrid crinoid cups are multi-plated structures of various shapes (Ausich 1988; Ausich et al. 1999), with plates united by rigid sutures (Ausich et al. 1999; Moore and Laudon 1943; Ubaghs 1978). Monobathrids are distinguished from the other order of camerate crinoids by the number of plate circllets at the base of the calyx below the arm rays. Monobathrids have only basals and radials. The remainder of the calyx in both orders is composed of several ranks of four basic types of plates: fixed brachials, anals, interrarial, and intrararial plates. Fixed brachial plates occur in each ray and are ranked from the primibrachial up to the highest rank included in the calyx. Between each ray are the interrarial plates; if they are absent the brachials directly contact each other. Intrararial plates occur between the fixed brachials of a single ray. The anal series occurs in posterior interray between rays C and D. Several ranges of anal plates may be present,

and this interray may be wider than other interrays. Maximum pentagonal symmetry is achieved in monobathrids with no anal series. The calyx contains the viscera of the crinoid. The calyx has two known functions: protecting the organs and providing a rigid base for the arms to operate (Ausich et al. 1999).

The organization of monobathrid calyx plates undergoes a distinct simplification over the history of the group. Early monobathrids from the Ordovician, e.g., *Reteocrinus*, have numerous fixed ray plates separated by a multitude of minute irregular interbrachial plates. At the simplest extreme are genera like *Platycrinites*, which has a calyx with only three unequal basals and five radials (Lane 1978). Simplification of the camerate calyx occurs by the upward displacement and subsequent elimination of all calyx plates above the radials (Moore and Laudon 1943).

No functional significance has been attributed to the reduction of plates over the Paleozoic in camerates (Simms 1990), but any number of circllets above two confers little extra rigidity at the stem-aboral cup interface (Ausich et al. 1999).

*Quantifying Calyx Complexity.*—Numerous metrics have been used for quantifying the complexity of biological structures (Cisne 1974; McShea 1991, 1992; Sidor 2001), but no single metric is appropriate for all purposes. For camerates, the simplification of the calyx is achieved by a reduction in both the number of plates and the number of types of plates. I gathered plate and type counts from published taxonomic descriptions, and all analyses are conducted on both of these measures.

Plate number (PN) is equal to the total number of plates in the calyx. Plates are counted only if they are restricted to the calyx. Plates that are also part of the tegmen are not counted. When a circllet is reduced from five to fewer by fusion, only the observed numbers of plates are counted. The number of plate types (NPT) is the number of named plate types that are present in the calyx.

*Stratigraphic Ranges of Genera.*—I use the genus ranges of Kammer and Ausich (2006), which are updated from the first and last occurrences of Sepkoski et al. (2002) for crinoid genera. First and last occurrences are

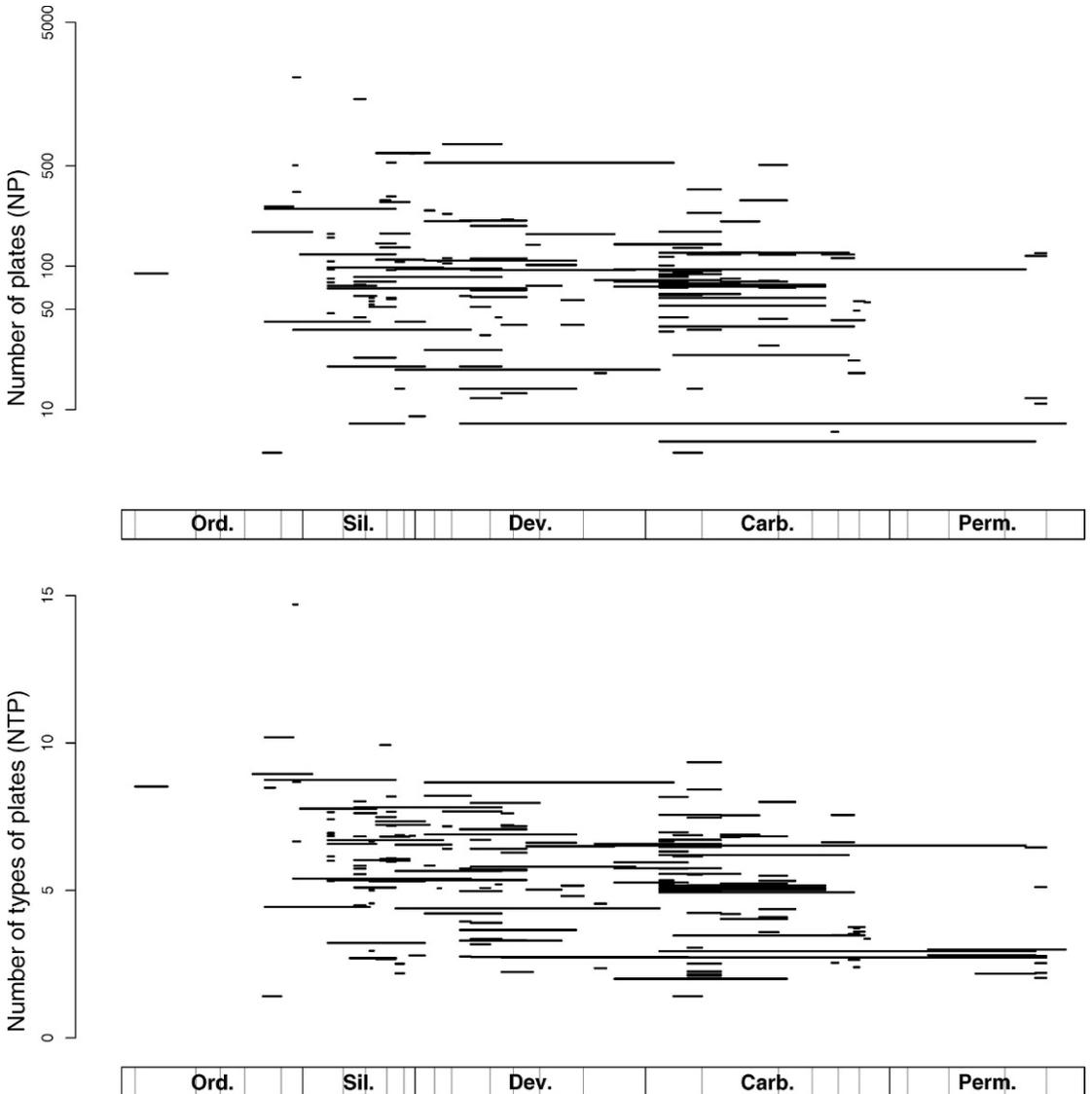


FIGURE 1. Two measures of genus-level calyx complexity are plotted over the Paleozoic. Each horizontal line is a genus stating at its time of first appearance and continuing until its time of last appearance. The top panel shows the number of plates (NP) for each genus of on a log scale. The bottom panel shows the approximate number of plate types (NTP) for each genus. A small amount of noise is added to the complexity metrics so that individual genera do not overlap on the plot. A total of 201 genera are included in this analysis.

resolved to the substage level whenever possible. Boundary ages for the time intervals are taken from Gradstein et al. (2004) for the post-Ordovician and Harland et al. (1990) for the Ordovician.

Genera not resolved to at least the stage level are not used for estimating taxonomic rates. For genera occurring in two or more intervals, the ages of the first and last

occurrences are taken to be the middle of the interval of first occurrence and the middle of the interval of last occurrence, respectively. Genera confined to a single interval are treated as ranging from the middle to the end of that interval. Figure 1 presents the stratigraphic ranges of genera sorted by their calyx complexity. A general decrease in complexity is observed.

## Methods

Because large-scale trends—by their very pervasiveness and persistence—can have complex dynamics, it has been conceptually useful to classify large-scale trends into those generated by passive mechanisms and those that are somehow driven or otherwise actively generated by selection (McShea 1994; Wagner 1996). According to McShea (1994) passive trend mechanisms have no inherent directionality, but directional patterns are due to structuring of the underlying state space. The archetypal passive mechanism was proposed by Stanley (1973). In Stanley's mechanism, random diffusion away from a morphological limit produces an increase in variance and a change in the mean phenotype. McShea (1994) also included species selection as a passive mechanism because he considered the relationship between diversification rate and morphology to be a characteristic of the state space. Driven mechanisms are processes that are inherently directional, where the direction of speciation is consistently biased regardless of causality (McShea 1994). The potential macroevolutionary trend mechanisms that are included in the driven trend class are (1) a bias in the direction of speciation caused by the constraints of heterochrony, (2) evolution along lines of least resistance, (3) microevolutionary or other systematic anagenetic change, and (4) directional differences in the magnitude of character state changes. Wagner (1996) proposed a class of active mechanisms that included driven ones as well as species selection and any other mechanism that would produce a systematic change in phenotypes over time.

Wang (2001) and Alroy (2000) have contributed to a more nuanced understanding of trend classes. Apparently passive trends can occur by a number of different nonlinear patterns of ancestor-descendant change (Alroy 2000). Wang developed a method to identify the ratio of passive to driven trend mechanisms. I take a different approach here that makes explicit the ways active, driven, and passive trend mechanisms contribute to generate a single trend.

Natural selection is a single process no matter what the units being selected are. So any single selection process can be described in terms of Price's Theorem (Frank 1997; Okasha 2007; Price 1972; Rice 2004). Price's Theorem describes the change in mean phenotype ( $\Delta\bar{\phi}$ ) over time as a function of both natural selection and any other processes that can change the mean phenotype across generations. The covariance between fitness ( $W$ ) and phenotypes ( $\Phi$ ) describes the change due to natural selection. Any changes in the mean phenotype that result from other processes are described in the second term ( $E(W\bar{\delta})$ ). Price's Theorem is (see Appendix for a derivation):

$$\Delta\bar{\phi} = \frac{1}{W} [\text{cov}(W, \phi) + E(W\bar{\delta})]. \quad (1)$$

When applied to large-scale trends, Price's Theorem aptly parses the change in mean phenotype to change due to species selection ( $\text{cov}(W, \phi)$ ) and change due to driven mechanisms ( $E(W\bar{\delta})$ ). I will make independent estimates of both of these terms from the fossil record.

*Species Selection.*—For any selection process, including species selection, directionality is a result of the selection differential. The selection differential ( $S_\phi$ ) measures the change in the mean phenotype due exclusively to species selection (where  $E(W\bar{\delta}) = 0$ ). Recall that the covariance of  $x$  and  $y$  can be rewritten as  $\beta_{x,y} \text{var}(\phi)$ , where  $\beta$  is the linear regression of  $y$  on  $x$  and  $\text{var}(\phi)$  is the variance of  $\phi$ . The selection differential is defined as

$$S_\phi = \text{cov}(W, \phi) = \beta_{\phi,W} \text{var}(\phi). \quad (2)$$

Here,  $\beta_{\phi,W}$  measures the linear regression between diversification rate ( $W$ ) and calyx complexity ( $\phi$ ). Much of the possible variation between fitness and phenotype has no bearing on the selection differential, as it is only the value of  $\beta_{\phi,W}$  that is incorporated into equation (2) (Rice 2004). This fact significantly simplifies interpreting the role of differential taxonomic rates in generating directionality in large-scale trends. Species selection occurs when the slope of the linear regression of taxonomic rate on calyx complexity is not zero. If the slope of linear regression is equal

to zero, species selection does not operate. How  $\beta_{\phi,W}$  is estimated will be discussed below.

Various methods have been proposed as tests for species selection (Gould and Eldredge 1977; Lieberman et al. 1993), but they suffer from not directly measuring diversification rates. In addition, species selection, like organismal selection, does not operate on individual lineages, but on the covariance between differential rates and phenotypes in a set of lineages, independent of the phylogenetic relationships between the lineages. Phylogenetic lineages may all have similar phenotypes distinct from other lineages, but that needn't be the case. In fact, for macroecological traits (e.g., geographic range) or general morphological traits (e.g., complexity or larval mode) unrelated taxa often have similar phenotypes.

The differential taxonomic rates necessary for species selection are often conceived of as a linear function of morphology (Gould 1990, 2002; Stanley 1975, 1979). Although a linear relationship need not be the case, the linear regression of fitness on phenotype is the quantity that determines the effects of *directional* selection on the mean phenotype (Rice 2004) and is the primary quantity that would generate a directional trend.

Unfortunately, the estimation of origination and extinction rates requires an ensemble of genera or species. This limits analyses to those that discretely approximate the relationship between phenotypes and differential rates by grouping phenotypes into categories, even if phenotypes vary continuously. For calyx complexity there is no natural way to group genera into morphological categories. Using a fixed set of morphological categories to estimate a time-series of differential rates can make detecting species selection difficult, because as the clade evolves it may shift out of a phenotypic category. The number of occupied categories would then decrease over time. The slope of the regression between differential rates and calyx complexity cannot be adequately estimated if the number of observations decreases over time. Alternatively, the morphological categories can be

dynamically redefined so that all are occupied during each time-slice. If the true relationship between fitness and phenotype is linear, then all approaches will estimate the selection differential accurately, because estimations of a linear slope are not sensitive to the distance between the points. Local estimates of the global linear regression are accurate. However, if the relationship between fitness and phenotype is nonlinear, estimates of the linear regression will be accurate only locally, but the mean of all local linear regressions accurately estimates the global linear regressions (Fig. 2).

I use two protocols for defining morphological groupings to estimate taxonomic rates for the quantitative plate counts (PN and NPT). Groups in each time interval are divided into quartiles dynamically defined for each interval and statically for the overall taxonomic group. For static quartiles, the quartile boundaries are constant over time, and quartiles may be empty occasionally, especially in later time-slices. The dynamic quartiles are defined on the basis of only those genera present in a specific cohort. The boundaries between quartiles always change from interval to interval, but the relative degree of complexity is maintained. For both types of quartiles definitions, the most complex is the fourth quartile, and the simplest is the first.

Per capita origination and extinction rates are calculated for each time interval using Foote's (2000) maximum likelihood metrics. This allows fluctuations in rates from interval to interval to be detected. Maximum likelihood extinction rates ( $\hat{q}$ ) are a function of the number of genera that pass through an interval ( $N_{bt}$ ), the number of genera that go extinct in that interval ( $N_b$ ), and the interval length ( $\Delta t$ ):

$$\hat{q} = -\log(N_{bt}/N_b)/\Delta t. \quad (3)$$

Origination rates ( $\hat{p}$ ) are a function of the number of genera originating in an interval ( $N_t$ ), the number of genera that pass through that interval ( $N_{bt}$ ), and the interval length ( $\Delta t$ ):

$$\hat{p} = -\log(N_{bt}/N_t)/\Delta t. \quad (4)$$

For each complexity metric, separate estimates of origination and extinction are made

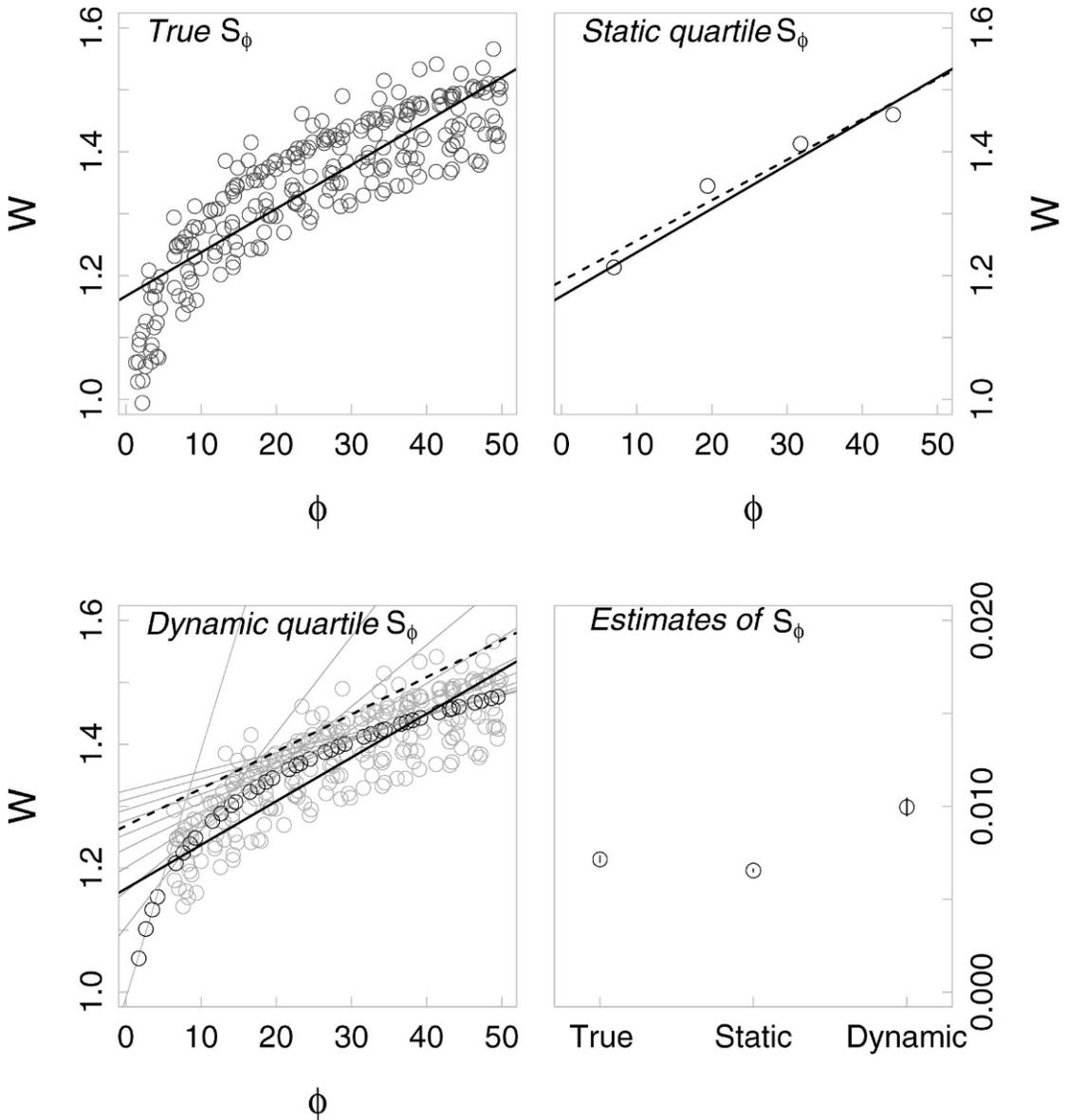


FIGURE 2. A hypothetical nonlinear relationship between phenotype ( $\phi$ ) and fitness ( $W$ ). When the mean phenotype shifts significantly over time, only a small portion of the full range of possible phenotypes are sampled at any time. The true selection differential is the linear regression through all the points (solid black lines). Estimates of the selection differential are shown as dashed lines. Two sampling protocols are used: static phenotypic quartiles and dynamic phenotypic quartiles. For dynamic quartiles, the selection differential can be estimated as the mean linear regression coefficient of temporal samples, even if the linear regressions vary in sign and slope (gray lines). The average estimate of the selection differential for static and dynamic protocols is compared with the true selection differential. Both sampling protocols recover the true selection differential with suitable precision.

for each of the dynamically and statically defined groups for each time interval. A single estimate of both origination rate and extinction rate is also made for the interval as a whole. A model selection approach (Burnham and Anderson 2002) is used to estimate

the empirical support for multiple rates or a single rate. Akaike's Information Criterion (AIC) (Akaike 1974) is a model-selection metric derived from information theory that measures the relative support, with respect to data, of a preselected set of models. Because

increasing the number of model parameters ( $K$ ) generally increases support ( $\log(L)$ ), AIC penalizes models by their complexity.

To factor in the effects of small sample sizes, I use  $AIC_c$  (Burnham and Anderson 2002):

$$AIC_c = -2 \log(L) + 2K + \frac{2K(K+1)}{n-K-1}. \quad (5)$$

Models with the lowest  $AIC_c$  are preferred. The AIC differences ( $\Delta_i$ ) values are used to rank models from highest to lowest confidence:  $\Delta_i = AIC_i - AIC_{min}$ . The preferred model is defined as:  $\Delta_i \equiv \Delta_{min} \equiv 0$ . For all  $n$  models,  $\Delta_i$  can be scaled to sum to one. These are the Akaike weights ( $\omega_i$ ):

$$\omega_i = \frac{\exp\left(-\frac{1}{2}\Delta AIC_i\right)}{\sum_{r=1}^n \exp\left(-\frac{1}{2}\Delta AIC_r\right)}. \quad (6)$$

The Akaike weights allow the set of models that fit the data to be selected following some criteria. Here, the presence of species selection is supported if the multi-rate model has larger AIC value, where  $\Delta_i$  for the multi-rate model is equal to 0. I used an Akaike weight of 0.89 for the cutoff of model preference. This is similar to using a likelihood criterion of rejecting hypotheses when an outcome is eight times less probable for one model than for the other (Wagner et al. 2006).

Species selection occurs when the Akaike weights for the multi-rate model are larger than 0.89 and the linear regression of diversification rate and calyx complexity is non-zero.

*Directional Bias.*—Driven mechanisms produce a non-zero value of the second term in Price's Theorem ( $E(W\bar{\delta})$ ). Phylogenetic mechanisms, such as a bias in the direction of speciation (McShea 1994), or directional variation in the magnitude of difference between ancestor and descendant morphologies (Wagner 1996), as well as phyletic mechanisms (e.g., microevolutionary change produced by a lower level of selection) all produce a non-zero value of this term. The value of  $\bar{\delta}$  can be directly estimated for phylogenetic mechanisms from a well-resolved phylogeny. Estimating  $\bar{\delta}$  for phyletic

change is more difficult, but in principle it is possible to do directly. Below, I will empirically estimate  $E(W\bar{\delta})$  with a new generalized form of McShea's (1994) subclade test. It is not possible with this indirect method to identify what driven mechanisms are occurring because it only estimates the strength of the directional bias, not how that bias is produced.

Driven trends are those where descendant species tend to be different from their ancestral species in a consistent way. McShea (1994) demonstrates that driven trends consistently produce subclades with a skewed morphological frequency distribution. He used this pattern to distinguish passive from weakly driven trends with the subclade test. Subclades originating distant from the bounding effects of a morphological constraint or limit will still be skewed if the trend is driven. Other methods of detecting driven trends, preferable to the subclade test but more data intensive, rely on robust phylogenies and known ancestor-descendant pairs (Alroy 2000; McShea 1994, 2000). Unfortunately, the phylogenetic relationships of monobathrid genera are not sufficiently resolved to use direct phylogenetic methods.

Because the subclade test was developed explicitly to distinguish between passive and weakly driven trends, McShea's method, as he used it, does not directly generalize to driven trends of all strengths. But the skewness of subclades can be used to estimate the strength of a directional bias indirectly. McShea's (1994) simulations show that the direction of skewness is a function of the direction of the underlying bias. A tendency for species to be larger than their ancestors translates to positive skewness; the tail of the subclade morphological frequency points in the direction of the bias. This is true when the bias is slight, and species are sampled at single time-slices.

I evaluate the generality of this conclusion by using a simple birth-death model where each new species is assigned a morphological value. In a run of the simulation, each species is assigned a duration, first appearance, and number of descendants based on an exponential, uniform, and Poisson distribution,

respectively. The morphology assigned to each species is a random uniform deviate added to or subtracted to the ancestral morphology, such that new species are different from their direct ancestor. The probability of increases can be varied by the directional bias. A bias of 0.5 is an unbiased random walk.

In addition to the frequency distribution constructed of members that cross a single time-slice, morphological frequency distributions can be constructed from all members of a subclade independent of when they lived. I simulated 1000 clades with a diversity of at least 50 members and estimated the skewness for the subclade as a whole and the skewness of the members occurring during the zenith of the subclade for nine values of directional bias.

The result for time-slice samples is strikingly different from what McShea (1994) found for weakly driven trends. He found that, in general, skewed tail points in the direction of the trend. However, his conclusion needs qualifying, because it is only true when the probability of change between ancestor and descendant is close to zero. When the probability that descendants differ in morphology from their ancestors is low, the majority of species in a subclade have identical morphologies and any time-slice skewness is directly a result of the few species that deviate morphologically. If the direction of speciation is unbiased, the expectation is equal to the number of species on either side of the mean morphology, yielding a symmetrical distribution. When the direction of speciation is biased, there will be a greater number of species on one side of the morphological frequency distribution. When the morphology of descendant species is different from its ancestor, the more numerous set of species will show a greater variance of morphology. This produces a sampled morphological frequency distribution that is skewed in the direction of the bias.

The time-slice skewness of subclades is directly opposite to the speciation bias when the probability of change is high. If there is no directional bias, the time-slice morphological frequency distribution of subclades will be

symmetrical: half of the descendants will be less than their ancestors and half will be greater in morphological value. Any bias changes the overall proportion of descendants that are greater than their ancestors. For example, if the bias is equivalent to 60 increases for every 40 decreases, the mode of the time-slice morphological frequency distribution will shift in the direction of the bias, producing a distribution skewed in the opposite direction of the bias.

The direction of skewness for morphological frequency distributions consisting of all members of the subclade depends on the strength of the directional bias. Strongly driven trends leave what amounts to stragglers behind the bulk of the species. The faster the subclade moves through morphospace the more distant the earliest members will lie in morphospace. This causes the direction of skewness to point in the opposite direction from the direction of the bias. Weakly driven trends send out few founders and the skewness points in the direction of the bias.

The strength of the directional bias can be indirectly measured from the direction of skewness of both whole and time-slice frequency distributions. The sign of the average skewness for both whole subclades and time-slice sample distributions has been calibrated to the directional bias by simulation. The calibration is presented in Table 1 and illustrated in Figure 3.

## Results

*Species Selection.*—Table 2 presents the AIC model selection results for each time interval. Diversification rates are estimated following three models separately for both NP and NTP measures of calyx complexity. First, a single estimate of origination rate and extinction rate is made for all genera combined. Diversification rate is calculated as origination minus extinction rate. Estimates of diversification rates for both dynamic and static quartiles are then made for each interval. Because static and dynamic morphological quartiles are different protocols for sampling the relationship between fitness and phenotype, they are not treated like competing models in the AIC analysis. Dynamic and static quartile models

TABLE 1. The direction of skewness for simulated morphological frequency distributions estimated from every member of a subclade and for those members that cross the time-slice with maximal diversity. A simple birth-death model (described in the text) for each bias parameter is used to correlate the bias parameter with the direction of skewness for both sampling protocols. One thousand simulation runs that produce a total diversity greater than 50 species were made for each parameter value.

Bias (probability of increase in complexity)	Skewness of all members	Skewness of members at the time of highest diversity
0.1	+	+
0.2	+	+
0.3	-	+
0.4	-	+
0.5	0	0
0.6	+	-
0.7	+	-
0.8	-	-
0.9	-	-

are compared separately with the single-rate model. AIC values for both complexity metrics have preferred support for species selection in each time interval during which a

comparison can be made. Table 3 compares the AIC values for the time-series of diversification rates for each model. Here, log-likelihoods are summed over the whole time-series and the resulting AIC values allow a ranking of the models I use to estimate the average diversification rate. Models of species selection are overwhelmingly preferred over a model where only a single diversification rate exists; the Akaike weights ( $\omega_i$ ) indicate no support for the single-rate model. Because models of species selection are preferred over the model with a single ensemble rate in all time intervals, the magnitude and direction of species selection can be measured.

A single-rate estimate is made for each phenotypic quartile in each time interval. The directional selection differential for an interval in time is a function of the linear regression of the diversification rate estimates on the median calyx complexity of each phenotypic quartile. The net effect of species selection can be measured by a single linear

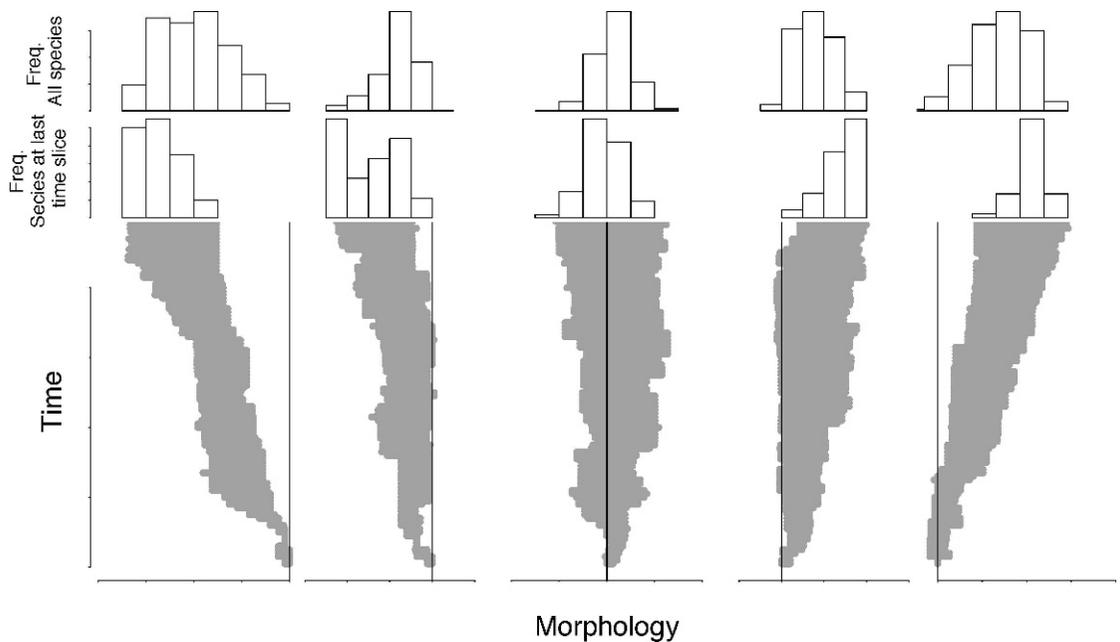


FIGURE 3. Simulated lineages illustrating how the generalized subclade test can estimate the strength of a driven trend. The generalized subclade test compares the skewness of the frequency distribution for all members of a lineage to the skewness of the frequency distribution of only those members occurring at a single time-slice. Five typical monophyletic lineages are shown in gray, each with a different probability of morphological change at the time of speciation. The vertical black line indicates the morphology of the first species in all lineages. Two sets of histograms are plotted for each lineage: the frequency distribution of species morphologies at the last time-slice, and the frequency distribution of all species morphologies in the lineage.



TABLE 3. A comparison of support for models of differential diversification rates and a for model of a single diversification rate. Log-likelihoods ( $\log L$ ) are summed across each parameter in the model. DAIC<sub>c</sub> show the difference between the small sample size AIC of the model minus the AIC for the best model. Akaike weights ( $\omega_i$ ) are also presented. Models that show greater support are indicated in bold.

Model	$\log L$	DAIC <sub>c</sub>	$\omega_i$
Single Rate NP	-269	281	0
<b>Static Quartile NP</b>	<b>-138</b>	<b>0</b>	<b>1</b>
Single Rate NTP	-269	300	0
<b>Dynamic Quartile NP</b>	<b>-90</b>	<b>0</b>	<b>1</b>
Single Rate NTP	-269	282	0
<b>Static Quartile NTP</b>	<b>-144</b>	<b>0</b>	<b>1</b>
Single Rate NTP	-269	390	0
<b>Dynamic Quartile NTP</b>	<b>-73</b>	<b>0</b>	<b>1</b>

regression between complexity and diversification rate or as the average of the per-interval linear regressions, even if the direction and magnitude change over time.

In Figure 4, I present the linear regressions estimated at each time interval for dynamic and static quartiles for both complexity measures (NP, NTP). Because the range of phenotypes changes over time, the average selection differential cannot be estimated by a single linear regression through the complexity-diversification rate data, which gives a slope opposite of the interval-to-interval slopes. Instead, I measure the net selection differential by taking the average slope of per-interval linear regressions.

There is no way to know a priori what the relationship between diversification rate and calyx complexity will be, so the static quartiles are used to estimate the global pattern and the dynamic quartiles estimate the local relationship. The average slope of both static and dynamic estimates is presented in the bottom panel of Figure 4. For both the number of plates (NP) and the number of types of plates (NTP) the average linear regression has a negative slope. This means that, on average, high calyx complexity is associated with negative diversification rate. Simple genera tend to have a positive diversification rate. This pattern in the covariation of diversification rates and complexity is consistent with the patterns of standing diversity seen in the fossil record of monobathrids because diversification rates

can be negative when diversity is high and positive when diversification is low. Early in their history, monobathrids are quite diverse and have a range of calyx complexities; however, these groups go extinct over the Paleozoic. Toward the end of the Paleozoic, the platycrinids are dominant but have comparatively low diversity, which is maintained by a low, but positive, diversification rate.

*Driven Mechanisms.*—Using families as subclades, I conducted a subclade test for camerate plate counts and plate types. Both total and time-slice subclade tests are presented in Figure 5. Because monobathrids, as a clade, originally had complex calyces but no known limit to complexity exists (a calyx could have as many plates and plate types as imaginable), there is no a priori expectation for the occurrence of passive diffusion away from a morphological boundary. However, passive diffusion can still be ruled out by plotting subclade skewness against the mean subclade complexity. If any unidentified boundaries exist they will be indicated by a correlation between skewness and mean complexity. Skewed subclades occur across the full range of complexity, so passive diffusion can be rejected as a mechanism. Subclades do vary in their direction of skewness. I use the sign of the average skewness for subclades as a whole and of time-slice subclades to estimate the directional bias and the expected change in mean morphology over time (Table 4). For the number of plates (NP) the mean skewness for subclades as a whole is 0.48, and the mean time-slice skewness for NP equals 0.20. For the number of types of plates (NTP) the whole subclade skewness equals 0.28 and has a time-slice skew of 0.17. Using Table 3, we can infer the average directional bias from the direction of both whole subclade and time-slice skewness. When both skewness measures are positive we can infer that there is a tendency for complexity to decrease due to driven mechanisms.

## Discussion

*Species Selection.*—Price's Theorem (eq. 1; also see Appendix) describes how the change

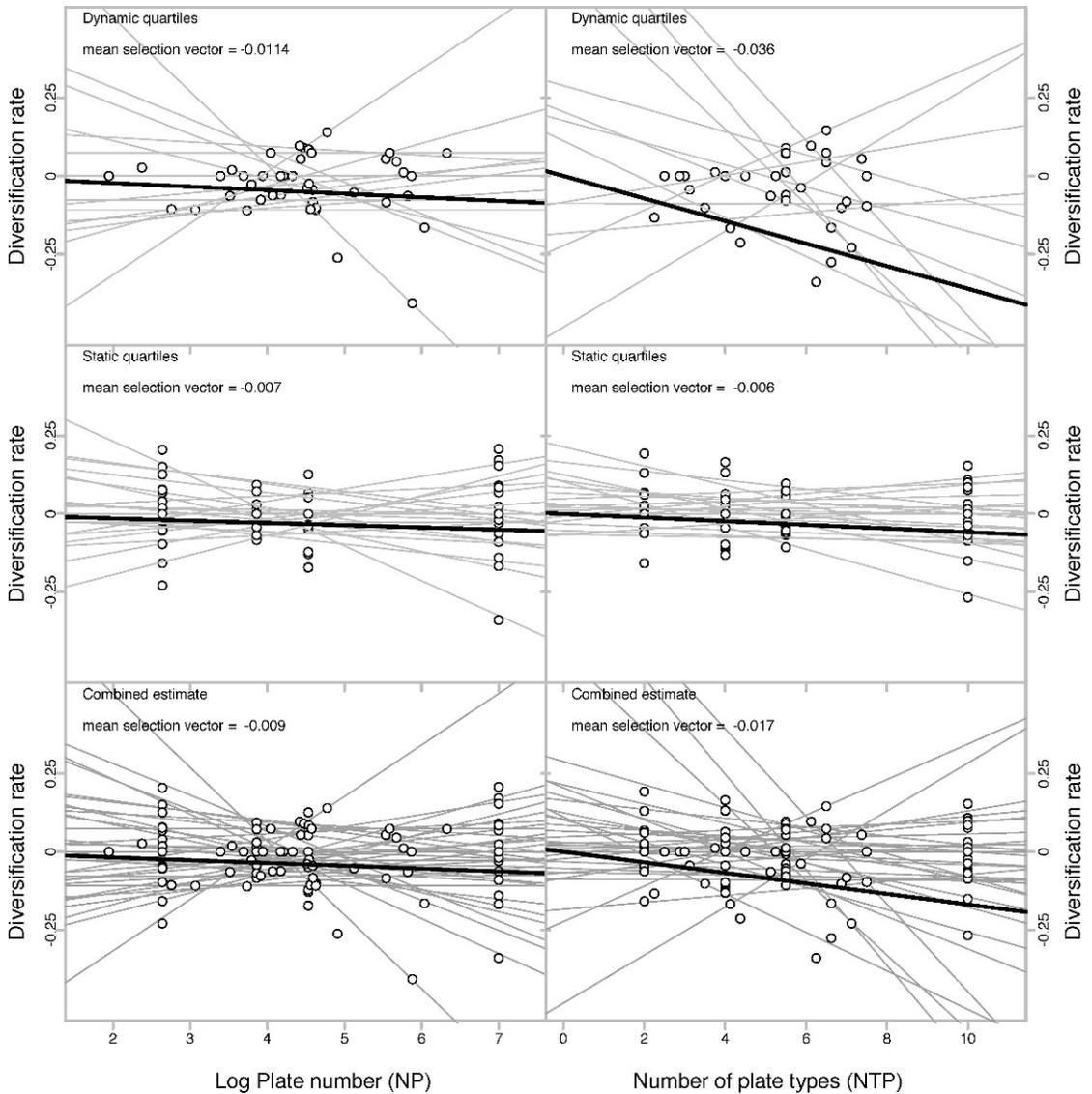


FIGURE 4. Estimates of the selection differential for monobathrid crinoids. Separate estimates are made for plate numbers (NP) and the number of plate types (NTP). Genera are grouped by their phenotypes into quartiles. Quartiles are either static, applied to each time interval in turn, or they are dynamic and redefined according to the range of phenotypes in that interval. Diversification rates are estimated for each quartile in an interval. The selection differential is the linear regression of diversification rate on the median phenotype of each quartile. The selection differential is estimated separately in each time interval (gray regression lines). Because of the variable direction of selection, the average of all slopes is used to estimate the overall selection differential for NP and NTP (lines plotted in black). Both static and dynamic quartiles are pooled in the bottom panels. The average slope of temporal estimates of both static and dynamic quartiles is the estimate of the selection differential for that trait.

in the mean phenotype is a function of both the change due to species selection and that due to driven mechanisms. This partition is possible because the effects of natural selection are easy to understand. The covariance between fitness and phenotype is precisely what we mean by natural selection, whatever

focal level we are studying. Furthermore, this covariance is independent of the phylogenetic relationships of the evolving units. All members of a phenotypic class have the same expected fitness independent of the possibility that their ancestors were of a different phenotypic class. For directional selection, the expected fitness is

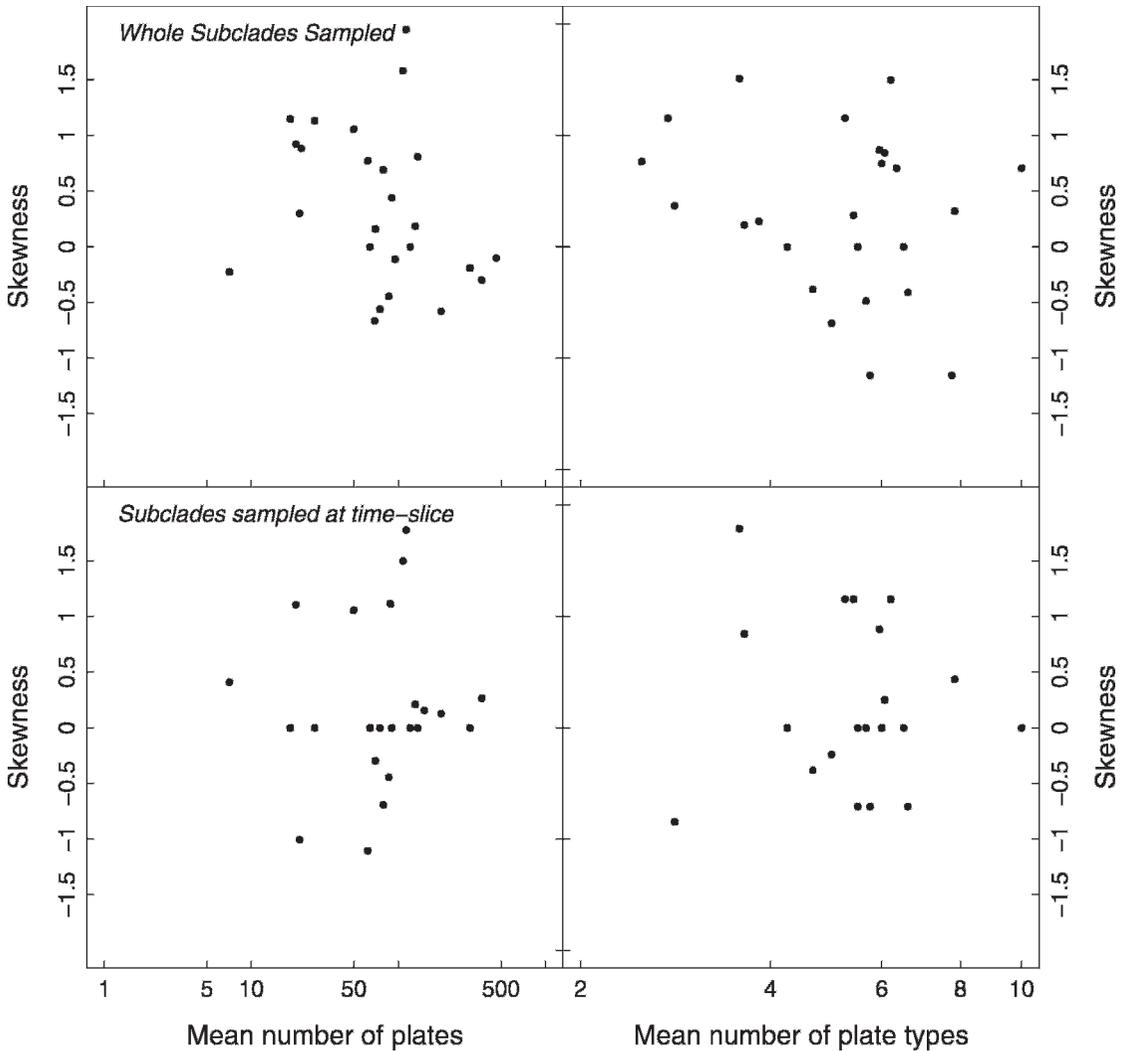


FIGURE 5. A subclade test of camerate families. There are two methods of sampling the skewness of subclades: (1) the skewness of the whole-subclade and (2) the skewness of the subclade at a single time-slice. Both are presented here; the top panels are the whole subclade test. In the bottom panels skewnesses are measured for each subclade at the time of its peak diversity. Skewness is plotted against the mean plate counts for each subclade. No correlation between plate count and skewness exists, indicating that no morphological boundaries are influencing skewness. The left panels show subclade tests for the total number of plates, while the right panels show subclade tests for the number of plate types.

simply the linear regression of estimated realized fitness on phenotype. The average linear regression is  $-0.009$  for NP and  $-0.017$  for NTP. These values, when multiplied by the pheno-

typic variance, give the net selection differential ( $S_\phi = \text{cov}(W, \phi) = \beta_{\phi, W} \text{var}(\phi)$ ;  $S_{NP} = -0.02077$ ;  $S_{NTP} = -0.1139$ ). This is the measurable net effect of species selection that is responsible, in part, for the simplification trend in monobathrids.

TABLE 4. Mean skewness of monobathrid families. Skewness for both protocols are positive for both complexity metrics. Using the calibration in Table 3, we see that the probability that a species will be more complex than its ancestor is less than 0.2.

	Mean subclade skew	Mean sampled skew
NP	0.48	0.2
NTP	0.28	0.17

A selection differential is estimated for each time interval from the diversification rates of the best-supported model. Recall that I compare two models of species selection, one where all phenotypes have the same rates and one where rates vary with phenotype. If the single-rate model has greater support, then

the estimated selection differential will always be equal to zero. If the multi-rate model has greater support, then the selection differential can have any value, including zero. Commonly linear regressions are used to estimate the causal relationship between two factors, and in these cases testing for the significance of the regression is important. Here, the significance of the linear regression is not important because the linear regression between phenotype and fitness itself determines the dynamics of evolution (Rice 2004).

The large variation in selection differentials from interval to interval makes it important to distinguish between the effects of species selection and its strength. The strength of species selection can be measured in each interval and can vary in magnitude, whereas the effect of species selection can be measured by the difference in morphology after some elapsed time due to species selection.

*Driven Mechanisms.*—The effects of all other processes that produce a directional change in phenotypes over time are partitioned into the second term of Price's Theorem. The classic driven mechanisms are all represented in Price's Theorem as the value of the second term. When  $E(W, \bar{\delta})$  is non-zero, there is a driven component. Similarly, passive diffusion away from a morphological boundary also produces a non-zero value in this term of Price's Theorem because the mean is expected to increase even if the minimum does not (McShea 1994). Because the average skewness of both sampling protocols are positive, the directional bias is estimated to be very strong; the probability of complexity increasing is at least 0.2. This means that at least 80% of genera can be inferred to be simpler than their ancestors. Unfortunately the indirectness of the subclade test does not allow the actual mechanism of directional change to be made. But because subclade skewness is not correlated with the mean subclade complexity, there is no effect of passive diffusion detectable. Bounded diffusion is the only passive mechanism that generates an expected direction change; thus the mechanism involved in monobathrids must be driven.

*Combining the Effects of Species Selection and Driven Mechanisms.*—In monobathrids, both

terms of Price's Theorem are estimated to be non-zero and negative. Species selection and driven mechanisms contribute in concert to the overall trend.

Traditionally, species selection is recognized if the null hypothesis of directionality produced by driven mechanisms and lower levels of selection is rejected (Gould and Eldredge 1977; Lieberman et al. 1993; Grantham 1995). Price's Theorem accurately models all selection processes, including macroevolutionary ones, and allows us to use a framework of estimation rather than null hypothesis testing in the study of species selection. The payoff of this approach is the ability to identify the contribution and interaction of different processes to a pattern when the processes produce roughly the same patterns. This approach is particularly relevant for macroevolution, where the role of multiple levels of selection and constraints in morphological evolution could interact in complex ways. Price's Theorem can be hierarchically expanded to describe multiple levels of selection (Hamilton 1975; Arnold and Fristrup 1982; Frank 1998; Okasha 2007) and driven change. To make the hierarchical expansion, note that the driven component of Price's Theorem,  $E(W, \bar{\delta})$ , contains the change in phenotype between ancestor and descendant ( $\bar{\delta}$ ). This is really just a change in phenotype over time and is itself a function of selection and driven mechanisms at a lower level, meaning that Price's Theorem also describes changes in  $\delta$ :

$$\delta_{upperlevel} = \frac{1}{\bar{W}} \left[ \text{cov}(W, \phi_{upperlevel}) + E(W\delta_{lowerlevel}) \right]. \quad (7)$$

Recursively expanding Price's Theorem allows it to describe multiple levels of selection. To add in a lower level of selection, equation (1) (this time indexed for individuals within a group) can be substituted for the mean change in phenotype:

$$\Delta \bar{\phi}_g = \frac{1}{\bar{W}_g} \left[ \text{cov}(W_g, \phi_g) + \frac{1}{N} \sum_g \left[ \frac{1}{\bar{W}_i} \text{cov}_g(W_i, \phi_i) + E_g(W_i \bar{\delta}_i) \right] \right] \quad (8)$$

What this means for interacting macroevolutionary processes is that, for example, a lower

level of selection and a bias in the direction of speciation both change the mean phenotype over time independent of selection at that level. Microevolution and focal-level driven mechanisms have the same effect on a trend, and both are independent of selection at the focal level. They both produce a systematic change in phenotypes over time, which is then sorted by selection. In other words, morphological stasis is not required for species selection to operate.

### Summary

The complexity of the monobathrid calyx is quantified as the number of plate types and the plate number. The complexity of the calyx decreases over time so that later camerates tend to have much simpler calyces.

Price's Theorem, an exact statistical description of the process of natural selection, is used to model the interaction of selection and driven trend mechanisms. The values of the terms of Price's Theorem are estimated from the fossil record of monobathrids. The vector of directional species selection is a function of the linear regression of diversification rate on calyx complexity. Differential diversification rates were identified by a model-selection approach and used to estimate the selection differential in each time interval. Genera with simple calyces tend to have a higher diversification rate than genera with complex calyces.

The strength of the driven mechanism was estimated by a generalized subclade test where the direction of skewness for subclades was calibrated to the strength of the directional bias by simulation. Monobathrid families were used as subclades, and the skewness of all genera in a family and the skewness of only those genera alive during the zenith of the family were calculated. A directional change, where about 80% of genera are simpler than their ancestors is indirectly estimated.

Both species selection and the driven component contribute to the overall simplification of the monobathrid calyx over time.

### Acknowledgments

I thank M. Bove, P. Harnik, L.H. Liow, and D. McCandlish for discussions, M. Foote, D.

Jablonski, L. H. Liow, D. McShea, L. Van Valen, and P. Wagner for reading versions of this manuscript, and W. Ausich and J. Finarelli for helpful reviews.

### Literature Cited

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19:716–723.
- Alroy, J. 2000. Understanding the dynamics of trends within evolving lineages. *Paleobiology* 26:319–329.
- Arnold, A. J., and K. Fristrup. 1982. The theory of evolution by natural selection: a hierarchical expansion. *Paleobiology* 8:113–129.
- Ausich, W. I. 1988. Evolutionary convergence and parallelism in crinoid calyx design. *Journal of Paleontology* 62:906–916.
- Ausich, W. I., C. E. Brett, H. Hess, and M. J. Simms. 1999. Crinoid form and function. Pp. 3–30 in H. Hess, W. I. Ausich, C. E. Brett, and M. J. Simms, eds. *Fossil crinoids*. Cambridge University Press, Cambridge.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York.
- Cisne, J. L. 1974. Evolution of the world fauna of aquatic free-living arthropods. *Evolution* 28:337–366.
- Foote, M. 2000. Origination and extinction components of taxonomic diversity: general problems. In D. H. Erwin and S. L. Wing, eds. *Deep time: Paleobiology's perspective*. *Paleobiology* 26(Suppl. to No. 4):74–102.
- Frank, S. A. 1997. The Price Equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* 51:1712–1729.
- . 1998. *Foundations of social evolution*. Princeton University Press, Princeton, N.J.
- Gould, S. J. 1990. Speciation and sorting as the source of evolutionary trends, or 'things are seldom what they seem.' Pp. 188–204 in K. J. McNamara, ed. *Evolutionary trends*. Belhaven, London.
- . 2002. *The structure of evolutionary theory*. Harvard University Press, Cambridge.
- Gould, S. J., and N. Eldredge. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3:115–151.
- Gradstein, F. M., J. G. Ogg, and A. G. Smith. 2004. *A geological time scale 2004*. Cambridge University Press, Cambridge.
- Grantham, T. A. 1995. Hierarchical approaches to macroevolution: recent work on species selection and the "effect hypothesis." *Annual Review of Ecology and Systematics* 26:301–321.
- Hamilton, W. D. 1975. Innate social aptitudes of man: an approach from evolutionary genetics. Pp. 133–155 in R. Fox, ed. *Biosocial anthropology*. Wiley, New York.
- Harland, W. B., R. L. Armstrong, A. V. Cox, L. E. Craig, A. G. Smith, and D. G. Smith. 1990. *A geologic time scale 1989*. Cambridge University Press, Cambridge.
- Hunt, G., K. Roy, and D. Jablonski. 2005. Species-level heritability reaffirmed: a comment on "On the heritability of geographic range sizes." *American Naturalist* 166:129–135.
- Jablonski, D. 1986a. Evolutionary consequences of mass extinctions. Pp. 382–418 in D. M. Raup and D. Jablonski, eds. *Patterns and processes in the history of life*. Report of the Dahlem Workshop on Patterns and Processes in the History of Life, Berlin 1985, June 16–21. Springer, Berlin.
- . 1986b. Larval ecology and macroevolution in marine invertebrates. *Bulletin of Marine Science* 39:565–587.
- . 1987. Heritability at the species level: analysis of geographic ranges of cretaceous mollusks. *Science* 238:360–363.

- Kammer, T. W., and W. I. Ausich. 2006. The "Age of Crinoids": a Mississippian biodiversity spike coincident with widespread carbonate ramps. *Palaos* 21:238–248.
- Lane, N. G. 1978. Family Platycrinitidae. Pp. T515–T516 in G. Ubaghs, et al. *Echinodermata 2 (Crinoidea)*. Part T (2) of R. C. Moore and C. Teichert, eds. *Treatise on invertebrate paleontology*. University of Kansas and the Geological Society of America, Lawrence, Kans.
- Lewontin, R. C. 1970. The units of selection. *Annual Review of Ecology and Systematics* 1:1–18.
- Lieberman, B. S., W. D. Allmon, and N. Eldredge. 1993. Levels of selection and macroevolutionary patterns in the turrillid gastropods. *Paleobiology* 19:205–215.
- Liow, L. H. 2006. Do deviants live longer? Morphology and longevity in trachyleberidid ostracodes. *Paleobiology* 32:55–69.
- McShea, D. W. 1991. Complexity and evolution: what everybody knows. *Biology and Philosophy* 6:303–324.
- . 1992. A metric for the study of evolutionary trends in the complexity of serial structures. *Biological Journal of the Linnean Society* 45:39–55.
- . 1994. Mechanisms of large-scale evolutionary trends. *Evolution* 48:1747–1763.
- . 2000. Trends, tools, and terminology. *Paleobiology* 26:330–333.
- Moore, R. C., and L. R. Laudon. 1943. *Evolution and classification of Paleozoic crinoids*. Geological Society of America Special Paper 46.
- Okasha, S. 2007. *Evolution and the levels of selection*. Oxford University Press, Oxford.
- Price, G. R. 1972. Extension of covariance selection mathematics. *Annals of Human Genetics* 35:485–490.
- Rice, S. H. 2004. *Evolutionary theory: mathematical and conceptual foundations*. Sinauer, Sunderland, Mass.
- Sepkoski, J. J., Jr., D. Jablonski, and M. Foote. 2002. *A compendium of fossil marine animal genera*. Paleontological Research Institution, Ithaca, N.Y.
- Sidor, C. A. 2001. Simplification as a trend in synapsid cranial evolution. *Evolution* 55:1419–1442.
- Simms, M. J. 1990. Crinoids. Pp. 188–204 in K. J. McNamara, ed. *Evolutionary trends*. Belhaven, London.
- Simpson, C., and P. Harnik. 2009. Assessing the role of abundance in marine bivalve extinction over the post-Paleozoic. *Paleobiology* 35:631–647.
- 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.
- . 1979. *Macroevolution*. W. H. Freeman, San Francisco.
- Ubaghs, G. 1978. Evolution of camerate crinoids. Pp. T281–T292 in G. Ubaghs, et al. *Echinodermata 2 (Crinoidea)*. Part T (1) of R. C. Moore and C. Teichert, eds. *Treatise on invertebrate paleontology*. University of Kansas and the Geological Society of America, Lawrence, Kans.
- Van Valen, L. 1975. Group selection, sex, and fossils. *Evolution* 29:87–94.
- Wagner, P. J. 1996. Contrasting the underlying patterns of active trends in morphologic evolution. *Evolution* 50:990–1007.
- Wagner, P. J., M. A. Kosnik, and S. Lidgard. 2006. Abundance distributions imply elevated complexity of post-Paleozoic marine ecosystems. *Science* 314:1289–1292.
- Wang, S. C. 2001. Quantifying passive and driven large-scale evolutionary trends. *Evolution* 55:849–858.

#### Appendix

*Price's Theorem.*—The following derivation of Price's Theorem follows Rice (2004). Let  $N$  be the population size,  $\phi_i$  be the phenotype of individual  $i$ , and  $\bar{\phi}$  be the mean phenotype. The phenotypic difference between the  $j$ th descendant of individual  $i$  is denoted  $\delta_{i,j}$ . The symbol for the difference between the mean descendant phenotype and  $\phi_i$  is  $\bar{\delta}_i$ . Individual  $i$  has  $W_i$  descendants.  $\bar{W}$  is the mean number of descendants per individual. Descendant  $j$ 's phenotype is equal to  $\phi_i + \delta_{i,j}$ . The mean phenotype of the descendants ( $\bar{\phi}$ ) is calculated by summing up the phenotypes of  $W_i$  descendants of individual  $i$ , repeating for the  $N$  phenotypes of the ancestral generation, and dividing by the total number of descendants:

$$\bar{\phi} = \frac{\sum_{i=1}^N \sum_{j=1}^{W_i} (\phi_i + \delta_{i,j})}{\sum_{i=1}^N W_i} \quad (A1)$$

Equation (1) can be rewritten as,

$$\bar{\phi} = \frac{1}{N\bar{W}} \left[ \sum_{i=1}^N W_i \phi_i + \sum_{i=1}^N W_i \bar{\delta}_i \right], \quad (A2)$$

which simplifies to

$$\bar{\phi} = \frac{1}{\bar{W}} [E(W_i \phi_i) + E(W_i \bar{\delta}_i)]. \quad (A3)$$

Recall that  $\text{cov}(x, y) = E(xy) - \bar{x}\bar{y}$  and substitute it for  $E(W\phi)$ :

$$\begin{aligned} \bar{\phi} &= \frac{1}{\bar{W}} [\text{cov}(W\phi) + \bar{W}\bar{\phi} + E(W_i \bar{\delta}_i)] \\ &= \frac{1}{\bar{W}} [\text{cov}(W\phi) + E(W_i \bar{\delta}_i)] + \bar{\phi}. \end{aligned} \quad (A4)$$

Subtracting  $\bar{\phi}$  from both sides gives the final form of Price's Theorem describing the change in mean phenotype over time ( $\Delta\phi$ ). This change is a function of the covariance between fitness ( $W$ ) and phenotype ( $\phi$ ) (the covariance represents the selection differential) and the expected mean phenotype after reproduction occurs ( $\bar{\delta}$  is the mean difference between ancestor and descendants). Price's Theorem is:

$$\Delta\bar{\phi} = \frac{1}{\bar{W}} [\text{cov}(W, \phi) + E(W\bar{\delta})]. \quad (A5)$$