Time, Species, and Separating Their Effects on Trait Variance in Clades

FOLMER BOKMA*

Department of Ecology and Environmental Science, Umeå University, Umeå S-90187, Sweden;
Correspondence to be sent to: Department of Ecology and Environmental Science, Umeå University, Umeå S-90187, Sweden;
E-mail: folmer.bokma@emg.umu.se.

Received 25 May 2009; reviews returned 24 August 2009; accepted 7 March 2010
Associate Editor: Todd Oakley

It remains largely unknown how phenotypic differences between species develop over time (Eldredge and Gould 1972; Gould and Eldredge 1993; Stanley 1998; Gould 2002; Reznick and Ricklefs 2009). From his early writings, it seems that Darwin initially envisioned macroevolution (evolution on a geological timescale) as the long-term consequence of microevolutionary changes in subsequent generations within populations. Darwin’s view of gradual, anagenetic evolution has been widely adopted by evolutionary biologists although paleontologists have repeatedly pointed out that several fossil records suggest a different pattern; of rapid evolution when a species appears, followed by morphological stasis during its subsequent lifetime (Eldredge and Gould 1972; Gould and Eldredge 1993; Stanley 1998).

In an attempt to estimate tempo and mode of phenotypic evolution for a large group of species, Ricklefs (2004) estimated by multiple regression the separate contributions of clade age and species number to phenotypic variance in subclades of passerine birds (Aves: Passeriformes). The idea of this statistical exercise was that the multiple regression would partition the morphological variance into components uniquely related to each of the independent variables (clade age and species number) and a component associated with the correlation between age and species number. Because species number but not clade age had a unique statistical influence on morphological variance, Ricklefs (2004) initially concluded that diversification was associated with speciation events.

That conclusion was disputed by Purvis (2004), who pointed out that “Under gradual change, variance accumulates along phylogenetic branches. Larger clades have more total branch length within them, even in same-aged clades, and so have more variance in gradually evolved traits.” Hence, one could not distinguish the effects of species number and time on morphological variance using Ricklefs’ (2004) approach. Subsequently, Ricklefs (2006) studied in detail how clade age and species number together determine the expected phenotypic variance in a clade and concluded that one cannot separate their contributions by multiple regression and therefore that his earlier conclusion that trait change was speciation dependent was not supported.
The considerations by Ricklefs and Purvis concerning "time, species, and the generation of trait variance in clades" are restricted to the expected (or mean) morphological variance in a clade. Here, I show that quantitative estimates of the rates of anagenetic and cladogenetic phenotypic evolution could be obtained using more sophisticated estimation methods that consider not only the mean variance but also the distribution of variances. In other words, the observation that species number predicts morphological variance in a clade better than clade age does, regardless of whether time or speciation events drive phenotypic differentiation, does not mean that one cannot separate these factors' contributions to phenotypic evolution.

**Model of Phenotypic Evolution**

Phenotypic evolution will be modeled here as a modified branching Brownian motion (Bokma 2002; Ricklefs 2006). Character evolution is often modeled as Brownian motion of the (logarithmic) trait value on the branches of a phylogeny, where the rate of anagenetic evolution is the Brownian variance, $s_a^2$. In the present model, species are also subject to cladogenetic change: at the instant a new species emerges, its trait value is different from that of its parent by a normally distributed amount. The rate of cladogenetic evolution is the variance of that (zero mean) distribution, $s_c^2$. If $s_c^2 = 0$, evolution is purely gradual. The phylogeny itself is assumed to result from a constant-rate, stochastic birth–death process with speciation and extinction rates $\lambda$ and $\mu$, respectively. Thus, a branch of a reconstructed phylogeny may contain several instances of cladogenetic evolution if the resulting species went extinct before the present (Bokma 2002), but all evolution that does not take place during speciation is considered anagenetic, even if it in reality were concentrated in short bursts interrupting stasis.

**Means and Distributions**

To begin with, I will give an example of the rather obvious fact that the interspecific phenotypic variance of a clade can result from purely gradual evolution as well as from purely speciational phenotypic evolution. I assumed arbitrarily $\lambda = 0.25$ and $\mu = 0.2$ per Myr and simulated 1000 phylogenies containing 25 species after 25 Myr (by disregarding simulations that led to a different number of species). On these simulated phylogenies, I then simulated purely gradual (i.e., $s_c^2 = 0$) phenotypic evolution, with a rate of $s_a^2 = 0.01$. Subsequently, I assumed that evolution is purely speciational (i.e., $s_c^2 = 0$) and simulated with $s_c^2 = 0.055$ because this led to approximately identical interspecific phenotypic variances: averaged over the 1000 phylogenies, $s_a^2 = 0.01$ and $s_c^2 = 0.055$ yield interspecific phenotypic variances $\overline{v}_a = 0.0978$ and $\overline{v}_c = 0.0971$, respectively (Fig. 1; unequal variance t-test, $P = 0.83$). This simple exercise merely confirms that observing the phenotypes of a collection of species at one particular point in time (e.g., the present) tells us nothing about the mode of evolution if we know nothing more about their evolutionary history than the time of their most recent common ancestor.

Interestingly, however, the distributions of simulated variances around the identical means are very different for $s_a^2 = 0.01$ and $s_c^2 = 0.055$ (Kolmogorov–Smirnov test $P < 0.0001$). As shown in Figure 1, the distribution of simulated variances is narrower when $s_c^2 = 0.055$ than it is when $s_c^2 = 0.01$. This implies that if we observe phenotypic variation of species from more than a single clade and if we assume that $s_a^2$ and $s_c^2$ are equal in these clades, then some combinations $[s_a^2, s_c^2]$ are more plausible as an explanation than other combinations.

**Estimation Using Artificial Neural Networks**

As a proof of the above allegation that rates of anagenetic and cladogenetic evolution can be estimated from the phenotypic variances of clades, I trained an artificial neural network for this purpose. This method, though rather unconventional, is conceptually more straightforward than mainstream estimation techniques. Elsewhere (Bokma 2006) I described in detail how one can use neural networks to estimate parameters, so I will keep methodological descriptions short here, focusing instead on the illustrative importance. Readers unfamiliar with neural networks are encouraged to read the detailed explanation (Bokma 2006).

For this demonstration, I used the ages and species numbers of the 95 passerine clades analyzed by Ricklefs...
I randomly sampled a combination of the parameters to be estimated (from uniform priors $0 < c^2_a < 0.02$ and $0 < c^2_c < 0.2$) and simulated phenotypic evolution for the 95 clades following the model explained above. Because we do not have species-level phylogenies for the clades, I also randomly constructed the phylogenies on which to simulate phenotypic evolution, using $\lambda = 0.1$ and $\mu = 0.07$. Thus, I assume here that the speciation and extinction rates are known so that the only parameters to be estimated are $s^2_a$ and $s^2_c$. That is of course not really the case (as discussed below) but greatly simplifies the estimation for the present illustrative purpose. Simulation was repeated for 10,000 random combinations ($s^2_a$, $s^2_c$). For every simulation (i.e., for every random combination), I calculated the cumulative sum of simulated phenotypic variances over the 95 clades. This cumulative sum can be thought of as a function increasing over the taxon numbers 1–95, the shape of which depends on $s^2_a$ and $s^2_c$.

Next, I trained an artificial neural network to associate the combination ($s^2_a$, $s^2_c$) used for simulation with the shape of the resulting cumulative sum of variances. Because the cumulative sum of variances of 95 clades is highly redundant, I presented the network with the first 10 principle components, which comprise close to all the variation in the data. The network consisted of 10 input neurons (for the 10 principle components), 25 neurons in a single hidden layer, and 2 output neurons (for $s^2_a$ and $s^2_c$), all with log-sigmoid transfer functions. The network was trained using 6000 simulations, 2000 simulations were used to stop training before the network started to overfit the data (Bokma 2006), and after training 2000 simulations were used to evaluate the network’s capacity to predict $s^2_a$ and $s^2_c$.

The predictions of the neural network are shown in Figure 2: clearly, the phenotypic variances of clades provide information about rates of anagenetic and cladogenetic phenotypic evolution because the network predicts both $s^2_a$ and $s^2_c$ independently. The estimation appears to be somewhat biased with a tendency to overestimate low rates of evolution and underestimate high rates, and estimation is not particularly precise either (Fig. 2). However, it is quite possible that the neural network is not a sufficient estimator: it does not necessarily use all the information in the data to estimate $s^2_a$ and $s^2_c$. Presenting the data to the network in a different way or training a larger network might also improve estimation. The purpose of the present exercise is not, however, to obtain precise estimates of $s^2_a$ and $s^2_c$ for simulated data sets. Instead, the aim was to show that phenotypic variances of clades can be used to estimate rates of gradual and speciational evolution. If that would not be possible, the neural network could not possibly have done it because it is free of any analytical solutions that potentially bias its performance. The neural network is merely a naive constellation of neurons that is presented with example data and turns out to be able to detect features in the phenotypic variances of clades that are associated with the rates of gradual and speciational evolution that generated these variances.

**Likelihood and Bayesian Estimation**

Now that we have seen that it is possible in principle to estimate the contributions of time and speciation to interspecific phenotypic variance, it is interesting to develop algorithms for accurate estimation. Obviously, the neural network estimation above was not particularly accurate, and even if it were, one might prefer conventional estimation techniques to facilitate hypothesis testing. Below, I will outline how Bayesian estimates of the rates of anagenetic and cladogenetic evolution could (theoretically) be obtained.

Again, we are interested in the rates of gradual ($s^2_a$) and cladogenetic ($s^2_c$) evolution, which we want to estimate from the interspecific phenotypic variances $v$ of clades with different ages $h$ and numbers of species $s$. The contribution of $s^2_c$ to $v$ depends, of course, on how...
frequently speciation takes place. If extinction never occurs, the number of speciation events in a clade is always one less than \( s \), but because extinction rates typically are not zero, \( v \) is influenced by the speciation and extinction rates \( \lambda \) and \( \mu \), respectively. It has been shown previously that \( \lambda \) and \( \mu \) can be inferred from \( h \) and \( s \) because (Bailey 1964; see also Bokma 2006)

\[
P(s) = (1 - \beta)\beta^{s-1}, \quad \text{where} \quad \beta = \frac{\lambda e^{(\lambda - \mu)h} - 1}{\lambda e^{(\lambda - \mu)h} - \mu}
\]

and where \( s > 0 \) because otherwise we would not be looking at the clade. That leaves only \( s_2^a \) and \( s_2^c \) to be estimated in the distribution \( \varphi(v|\lambda, \mu, h, s, s_2^a, s_2^c) \), so that it would be possible to estimate \( s_2^a \) and \( s_2^c \) using the likelihood

\[
\prod_{i=1}^{n} \varphi(v_i|\lambda, \mu, h_i, s_i, s_2^a, s_2^c),
\]

where \( n \) is the number of clades for which data are available. (In this particular notation, the likelihood assumes that \( \lambda, \mu, s_2^a, \) and \( s_2^c \) are equal across clades.)

The most convenient way to simultaneously infer \( s_2^a, s_2^c, \lambda, \) and \( \mu \) is probably by Bayesian inference, constructing a Monte Carlo Markov Chain (MCMC) using Metropolis–Hastings sampling (Metropolis et al. 1953; Hastings 1970). The structure of the algorithm would then be the following:

1. Assign initial values to \( s_2^a, s_2^c, \lambda, \) and \( \mu \) and evaluate the likelihood using Equation 2 (I assigned, largely arbitrarily, 0.01, 0.1, 0.5, and 0.47, respectively.)
2. Update \( \lambda \) and \( \mu \) using Equation 1
3. Update \( s_2^a \) and \( s_2^c \) using Equation 2
4. Repeat Steps 2 and 3 many times.

It should be noted that using Equation 2 and this algorithm structure, \( \lambda \) and \( \mu \) are assumed equal across clades, and the phenotypic data \( v \) have no bearing on the estimates of \( \lambda \) and \( \mu \). That is an arbitrary choice, but alternative assumptions do not change the main point of this contribution: that we can obtain estimates of \( s_2^a \) and \( s_2^c \) from observations of \( h, s, \) and \( v \).

The obvious problem with the above algorithm is that we do not know the analytical form of Equation 1, that is, \( \phi \) is unknown. It is well known that the particles at the front of a branching Brownian motion are approximately normally distributed. The variances of independent normally distributed random variables can be shown to be \( \chi^2 \) distributed. However, species are not statistically independent due to shared ancestry, and consequently the phenotypes of the tip species are not independent draws from a normal distribution (unless the trait under study shows no phylogenetic inertia). Hence, the phenotypic variance of species in a clade is only approximately gamma distributed, and I am not aware of a parametric family of statistical distributions that describes the distribution of the phenotypic variance.

### APPROXIMATE BAYESIAN COMPUTATION

A recent technique to estimate parameters without a likelihood function is approximate Bayesian computation (ABC; Marjoram et al. 2003). In conventional Bayesian estimation using MCMC, the likelihood is used to decide whether proposed parameter values are “plausible” enough to be accepted as a sample from the posterior distribution. ABC uses simulated data instead of the likelihood: How well simulated data resemble observed data depends on the parameter values (here \( s_2^a \) and \( s_2^c \) used for simulation, so that plausible parameter values can be identified comparing simulated with observed data. Thus, in ABC the proposed parameters are used to simulate data, after which the proposal is accepted if the simulated data closely resemble the observed data. For the present case of estimating rates of gradual and cladogenetic evolution, we could use the following algorithm:

1. Sample \( (\lambda, \mu, s_2^a, s_2^c) \) from their priors
2. Using observed ages and species numbers of clades, simulate phenotypic evolution on randomly constructed phylogenies
3. Calculate the distance between observed and simulated interspecific variances
4. Repeat Steps 1–3 many times.

As a measure of distance between observed and expected variances, I simply used the sum of squared errors after square root transformation:

\[
d = \sum_{i=1}^{n} (\sqrt{v_{\text{obs},i}} - \sqrt{v_{\text{sim},i}})^2,
\]

where the summation is over clades. Crucial in ABC is to decide the critical distance \( d_{\text{crit}} \) between observed and simulated data: If that distance is too small (e.g., \( d_{\text{crit}} = 0 \)), we will sample the exact posterior distribution, but proposals will rarely be accepted and the analysis cannot be completed in reasonable time. If on the other hand, \( d_{\text{crit}} \) is too large, all proposals will be accepted so that instead of sampling from the posterior distribution we sample from the prior.

I applied ABC to estimate rates of anagenetic and cladogenetic evolution \( s_2^a \) and \( s_2^c \) from Ricklefs’ (2004) passerine clades, using the interspecific variances of the first principle component only. (The first component explained 75% of morphological variance.) I assumed Dirac delta priors \( \lambda = 0.2 \) and \( \mu = 0.08 \), so that only \( s_2^a \) and \( s_2^c \) have to be estimated. After some initial trial and error, I decided on \( d_{\text{crit}} = 2 \).

Estimates of rates of anagenetic and cladogenetic evolution are shown in Figure 3. If it is assumed that evolution is purely gradual (i.e., \( s_2^a = 0 \)), the estimate of the rate of anagenetic evolution is \( s_2^c = 0.0118 \). If both rates are estimated, then the estimate of the rate of anagenetic evolution becomes lower (\( s_2^a = 0.0078 \)), as part of the interspecific phenotypic variance is accounted for by cladogenetic evolution (\( s_2^c = 0.0398 \), see also Fig. 3).
can easily calculate how much of the interspecific phenotypic variance is explained by anagenetic evolution as 0.0078 / 0.0118 = 66% (Mattila and Bokma 2008). The remaining 34% of phenotypic differences between species is due to rapid change during or immediately upon speciation.

Finally, we may want to determine which model is supported by the data: purely gradual evolution or a combination of anagenetic and cladogenetic evolution. In ABC, Bayes factor can (for low-dimensional estimation exercises like the present) be calculated from the acceptance rates of the 2 models. $s_2^2$ was proposed 242 times to obtain 100 samples in the purely gradual model, whereas in the 2-parameter model 8568 proposals were needed to obtain 1000 samples. Thus, we obtain a Bayes factor of 3.5 in substantial support of the purely gradual model.

**DISCUSSION**

Whether the approaches illustrated here are of practical value depends on a number of biological questions. For example, the current analysis assumes that phenotypes of species evolve largely independent of the phenotypes of contemporary species. Another assumption is that speciation and extinction rates have been constant over time and across species. Recent evidence suggests that both speciation and extinction rates are higher in large-bodied species (Liow et al. 2008). In addition, speciation rates may depend on the number of coexisting species (Rabosky and Lovette 2008). A similar, even more important assumption is that the rates of phenotypic evolution have been constant across the analyzed clades and over time. There is some evidence that this is not the case for birds (Bokma 2004; see also Monroe and Bokma 2009). If $s_2^2$ and $s_2^c$ are assumed to differ between clades in an idiosyncratic fashion, it is impossible to estimate them using any of the techniques described here. It is also quite possible that speciation rates depend not only on the number of coexisting species but also on the phenotypic similarity of the coexisting species (McPeek 2008). In species-poor communities with much phenotypic variation, speciation may be more likely, and phenotypic evolution more rapid, than in communities densely packed with phenotypically similar species.

It would certainly be interesting to know whether emergence and extinction of species depend on their (parents') phenotypes or whether the same processes drive both speciation and phenotypic evolution. If the latter is true, it will be very hard to estimate rates of gradual and speciational phenotypic evolution from the interspecific variances of clades. However, in order to test hypotheses that distinguish between the above scenarios, one has to start with the simplest model, and therefore it remains interesting to develop methods for estimation of rates of speciational evolution by comparing phenotypic variances of clades.

There exist alternative methods to estimate rates of anagenetic and cladogenetic phenotypic evolution, which can be used when a species-level phylogeny is available (Bokma 2008). Species-level phylogenies provide more information about the course of phenotypic evolution, and consequently, these methods must be expected to lead to better estimates. In the first place, the information the phylogeny provides allows for more accurate estimates of speciation and extinction rates, which are essential to estimate the contribution of speciation to phenotypic evolution. Second, a phylogeny provides the information needed to study phenotypic changes in ancestor-descendant pairs, which greatly increases power to infer tempo and mode of character evolution.

Nevertheless, increased statistical power will not be the most important advantage of species-level over higher-level phylogenies to estimate the mode of evolution. Interspecific variances of tens of clades are sufficient to estimate just 2 rates of evolution with reasonable accuracy (Fig. 3). Hence, the main advantage of species-level phylogenies is the increased power to investigate alternative models, for example, with variable speciation and extinction rates (Rabosky and Lovette 2008) or variable rates of phenotypic evolution (Monroe and Bokma 2009).

As compared with a higher-level phylogeny, a species-level phylogeny of the same group of species generally provides much more information about the course of phenotypic evolution. However, exactly how well tempo and mode of character evolution can be inferred depends only partly on phylogenetic resolution. What is most important has not changed from the earliest attempts to use molecular distances to reveal the mode of character evolution (Avise and Ayala 1975; Avise 1977): the possibility to compare very recent with very old groups—either sister species or clades.
FUNDING

This study was supported by the Swedish Research Council and Umeå University. The author awards SEK 10,000 to the first who provides an analytical form of $\phi$.

REFERENCES
