

**Biol 206/306 – Advanced Biostatistics**  
**Lab 11 – Models of Trait Evolution**  
**Fall 2016**

**By Philip J. Bergmann**

**0. Laboratory Objectives**

1. Explore how evolutionary trait modeling can reveal different information from evolutionary correlation and regression
2. Learn how the evolution of continuous traits can be modeled
3. Learn how to fit trait evolution models to data with phylogenies
4. Learn how to test evolutionary hypotheses using modeling

**1. Modeling Trait Evolution**

In the last lab we learned how to do evolutionary correlation and regression analyses by estimating evolutionary VCV matrices and doing PGLS. *What are the biological hypotheses being tested with these two approaches?*

Therefore, these techniques are best at addressing a relatively narrow range of evolutionary questions. Another approach that has been taken, that is broader, is using models to test for various biological phenomena, or simply to describe the evolution of a continuous trait (note that discrete traits can and are also modeled, but we will not be studying those models in this course due to time limitations). Trait evolution phenomena such as adaptive radiation, species specialization, and punctuational change can all be tested using well-implemented trait evolution models. One can also quantify the rate of evolution of different traits, or a single trait in different clades. Other aspects of trait evolution, such as the co-evolution of multiple traits and directional evolution have also been modeled. We will learn about many (but not all) of these models today.

Last week, you learned that evolutionary correlation and regression analyses assume a Brownian Motion (BM) model of trait evolution. The important foundation for this week is that BM can be used to model trait evolution. Actually, this model is the basis of more complicated models that we will explore later in today's lab. What makes the BM model attractive is its simplicity, and so it can be considered a null model, where a trait simply evolves in random directions during each generation or time interval. This may at first seem silly, but consider that environmental fluctuations can often be near-random, resulting in fluctuating selective pressures on the trait you are studying. Factors such as genetic drift and random mating can also result in fluctuating, random evolution of a trait.

The random direction of evolution component of the BM model results in a null expectation of no change in trait mean if you simulate trait evolution many times (see the figure illustrating BM

trait evolution through time in lecture). The variance around zero increases with time (again, when simulated many times). The variance, however, can increase at different rates, in proportion to how quickly the trait evolves. This rate of evolution is modeled as the  $\sigma^2$  parameter of the BM model. If you fit the BM model to different traits, or to the same trait for different clades, you can quantify and compare rates of evolution. *What biological hypothesis could you test by fitting a BM model to several traits in a clade of taxa?*

## 2. More Complex Models of Trait Evolution

Unsurprisingly, in some cases the BM model does not fit the data well. One reason for this may be that rates of evolution can change through time or for different clades. For this reason it is also useful to consider more complex models of trait evolution. Most, if not all of the models that have been implemented use the BM model as a starting point and add complexity. What this means is that these more complex models also have the evolutionary rate parameter ( $\sigma^2$ ), but then have additional parameters that are used to describe a trait's evolutionary history.

The details of the trait evolution models that we will use today are given in the accompanying lecture, but are briefly summarized here as well:

Model	K	Parameters	Notes
BM	2	$\sigma^2$	<ul style="list-style-type: none"> <li><math>\sigma^2</math> gives the rate of evolution of a trait for a clade</li> <li>This is the null model for trait evolution</li> </ul>
BM + $\lambda$	3	$\sigma^2, \lambda$	<ul style="list-style-type: none"> <li>Internal branches are multiplied by <math>\lambda</math></li> <li><math>\lambda</math> is a measure of phylogenetic signal, range: 0-1</li> <li><math>\lambda = 1</math>: Trait evolution follows the phylogeny, BM evolution</li> <li><math>\lambda = 0</math>: Trait evolves independently among lineages (do not need to take phylogeny into account)</li> </ul>
BM + $\kappa$	3	$\sigma^2, \kappa$	<ul style="list-style-type: none"> <li>All branch lengths are raised to the power <math>\kappa</math></li> <li><math>\kappa</math> can be used to test for punctuational evolution, range: <math>&gt;0</math></li> <li><math>\kappa = 1</math> indicates gradual BM evolution</li> <li><math>\kappa = 0</math> indicates punctuational evolution, where all change occurs at time of speciation</li> <li><math>0 &lt; \kappa &lt; 1</math>: More trait evolution than expected on shorter branches, so more stasis on longer branches</li> <li><math>\kappa &gt; 1</math>: More trait evolution than expected on longer branches</li> </ul>
BM + $\delta$	3	$\sigma^2, \delta$	<ul style="list-style-type: none"> <li>Node depths (distances from tips) are raised to the power <math>\delta</math></li> <li><math>\delta</math> can be used to test if trait evolution follows a pattern of adaptive radiation or species specialization, range: <math>&gt;0</math></li> <li><math>\delta = 1</math>: indicates gradual BM evolution</li> <li><math>0 &lt; \delta &lt; 1</math>: Most trait evolution is near the base of the tree</li> <li><math>\delta &gt; 1</math>: Most trait evolution occurs near the tips of the tree</li> </ul>

Early Burst	3	$\sigma^2_0, r$	<ul style="list-style-type: none"> <li>• Can be used to model adaptive radiation, where the rate of evolution slows down as niches are filled</li> <li>• <math>\sigma^2_0</math> is the initial rate of evolution at the base of the tree</li> <li>• <math>a</math> determines how <math>\sigma^2_0</math> changes through time</li> <li>• <math>r = 0</math>: trait follows pure BM evolution, <math>\sigma^2</math> is constant</li> <li>• <math>r &lt; 0</math>: <math>\sigma^2</math> decreases exponentially through time</li> <li>• <math>r &gt; 0</math>: increases exponentially through time</li> </ul>
Ornstein-Uhlenbeck	4	$\sigma^2, \theta, \alpha$	<ul style="list-style-type: none"> <li>• Is used to model persistent stabilizing selection around an optimal trait value</li> <li>• <math>\theta</math> is the trait optimum</li> <li>• <math>\alpha</math> is the strength of stabilizing selection</li> <li>• The higher <math>\alpha</math>, the stronger the stabilizing selection and the more constrained the rate of evolution relative to pure BM</li> <li>• If <math>\alpha = 0</math>, then the trait evolves following pure BM</li> <li>• If <math>\alpha</math> approaches zero and <math>\theta</math> is outside the range of taxon trait values, then the model suggests directional evolution</li> </ul>

One useful way of thinking about what the  $\lambda$ ,  $\kappa$ , and  $\delta$  parameters do is to compare the branch lengths on your original tree to the branch lengths on the tree that is transformed by these parameters. The original tree has branch lengths proportional to time, while the transformed tree has branch lengths proportional to the amount of evolution in your trait. If your trait follows pure BM, then the temporal and the trait evolution branch lengths are equivalent.

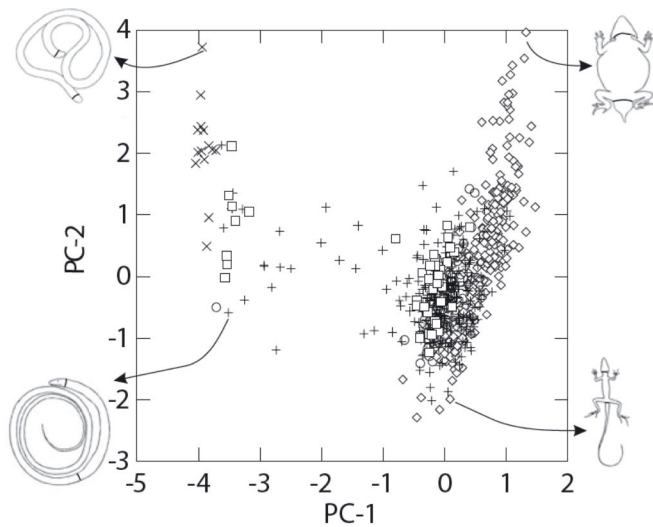
Also note that the BM +  $\delta$  and the Early Burst models are two alternative ways of testing for the same thing.

***For the BM +  $\delta$  model, what value of  $\delta$  would be consistent with adaptive radiation? Why?***

***What values of  $r$  in the Early Burst model would be consistent with adaptive radiation? Why?***

### **3. Fitting Models of Trait Evolution**

Today, you will be analyzing another lizard body shape dataset to characterize how certain traits have evolved. The dataset consists of two principal components from a PCA on body shape in 636 species of lizards, published by Bergmann & Irschick (2012. Vertebral evolution and the diversification of squamate reptiles. *Evolution* 66: 1044-1058). As can be seen from the figure



from the article (to the left), PC-1 represents body shape variation associated with the degree of elongation: low PC-1 values indicate elongate, limbless bodies, while high values indicate shorter, lizard-like bodies. PC-2 can be interpreted as an index of robustness: low values indicate gracile, thin bodies, while high values indicate robust, heavy, thick bodies. The dataset also contains  $\log(\text{Head Length})$ , which can be used as a measure of body size. The reason we use head length is because the body is a poor indication of general body size because some animals are

elongate. There is also a phylogeny that relates all of these species. **Load the package “geiger”, which will automatically load “ape”. Then load the dataset and the phylogeny into R, assigning the former to object “data” and the latter to “tree”. Check to ensure that the taxa in the dataset and on the phylogeny match exactly. Is the tree ultrametric and fully resolved?**

You will fit each of the models described above to each of the variables in the dataset to describe the evolution of each of the traits (PC-1, PC-2, LHL). You can do this handily using a single function that is part of the “geiger” package: **fitContinuous**. The function allows you to fit each of these models, and outputs the parameter values, the  $\ln(\text{likelihood})$ , AIC and AICc. Use the line:

```
> ? fitContinuous
```

To call a webpage that describes how to use the function. This is a very useful approach to learning how to use a new function. **Fit each of the models described above using the fitContinuous function and save the output for each run as an object.** Note that some of these models may take some time to fit, given the large size of the dataset and the complexity of the underlying mathematics. You may wish to divide up the models between you and your partner so that you can run two at a time. You can also encapsulate multiple model fittings within a set of  $\{\dots\}$ , which will then run them as a batch instead of singly. This saves time and can allow you to take a 15 minute break while the models are being fitted.

A few other notes about model fitting are needed. First, you can also fit a BM and OU model using the packages “OUCH” and “OUwie” (other packages will also fit BM models), but it is often inappropriate to compare models fit by different packages (unless you know exactly what you are doing) because different packages may use different algorithms to parameterize the models. This means that parameter values may be slightly different, and likelihoods can be very different, meaning that the models cannot be compared. The second important note is that the

“geiger” does not estimate a  $\theta$  parameter for the OU model, so the number of parameters is three (the output from “OUCH” does fit this parameter).

**Assignment: (10 points)**

*Complete the tables below for each of the traits. The “Parameter” column should be used to provide the value for the additional parameter that is estimated for some of the models. Fill in what the parameter is for each model. Then answer the questions.*

**PC-1: Degree of Elongation**

Model	K	$\sigma^2$	Parameter	$\ln(L)$	AIC <sub>c</sub>	$\Delta_i$	w <sub>i</sub>
BM							
BM+ $\lambda$							
BM+ $\kappa$							
BM+ $\delta$							
EB							
OU							

**PC-2: Degree of Robustness**

Model	K	$\sigma^2$	Parameter	$\ln(L)$	AIC <sub>c</sub>	$\Delta_i$	w <sub>i</sub>
BM							
BM+ $\lambda$							
BM+ $\kappa$							
BM+ $\delta$							
EB							
OU							

**LHL: Measure of Body Size**

Model	K	$\sigma^2$	Parameter	$\ln(L)$	AIC <sub>c</sub>	$\Delta_i$	w <sub>i</sub>
BM							
BM+ $\lambda$							
BM+ $\kappa$							
BM+ $\delta$							
EB							
OU							

*Which model fits the PC-1 data the best? What is your biological interpretation of how lizard body shape, as quantified by PC-1, has evolved?*

***Which model fits the PC-2 data the best? What is your biological interpretation of how lizard body shape, as quantified by PC-2, has evolved?***

***Which model fits the LHL data the best? What is your biological interpretation of how lizard body shape, as quantified by LHL, has evolved?***

***Both the  $BM+\delta$  and the EB models can be used to model trait evolution that we would expect under adaptive radiation. Do they seem to give consistent results in your analyses of PC-1? Explain your reasoning.***

***Which trait evolved the fastest and which evolved the slowest? Explain your reasoning.***

***For PC-1, why is the AICc for  $BM+\lambda$  higher than for BM? Why is this question being asked?***