Lecture notes
10/15 Brownian Motion & Continuous Trait Evolution

I. Random Walks - "The Drunkard's Walk"
   A. Pioneered in application to evolutionary biology by David Raup (paleontology) and Luca Cavalli-Sforza & Anthony Edwards (population genetics). Later by Russell Lande (Quantitative Genetics).
   B. In population genetics, a random walk is an approximation of allele frequency change under Brownian Motion. (Only when away from the boundary, breaks down because Brownian Motion does not have bounds or absorbing states!).
   C. Lande showed that Brownian Motion also occurs under genetic drift in complex, polygenic characters.
II. Brownian Motion - A continuous-time random walk
   A. Unbiased Brownian motion 2 parameters:
      1. σ^2 - Rate of evolution
      2. z_0 - The root state
   B. Brownian motion with a trend (or a biased random walk) has another parameter:
      1. μ - The trend parameter
   C. With extant only data (no fossils & an ultrametric tree) you CANNOT distinguish unbiased and BM with a trend models, there is no information about μ if all data is taken from the same time slice. Only with fossils are such models identifiable.
   D. A collection of lineages evolving under Brownian Motion will have an expected value of:

   $z_0 + \mu t$ -> expected mean of *N* lineages after t units of time

   $\sigma^2 t$ -> expected variance among *N* lineages after t units of time

   Variance thus increases linearly with time and Brownian motion is unbounded. If the BM is unbiased, then the expectation for all lineages is simply the root state.

III. Brownian Motion on a phylogeny
   A. Brownian Motion on a phylogeny results in species *covarying* with each other proportional to the amount of time they spent as the same lineage. Thus, we can construct a *Variance-Covariance Matrix* from the phylogeny that provides the expected covariances under a Brownian Motion model.
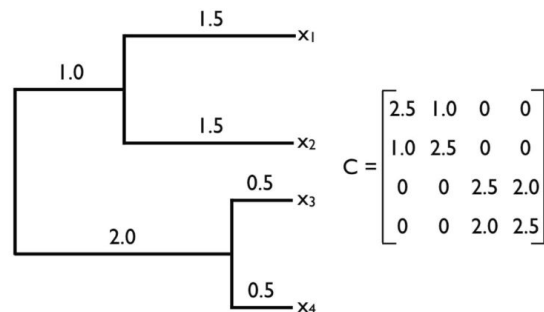
B. If two species share no evolutionary history (relative to the tree), they are completely independent, and their covariance is 0. In the tree above, species x1 and x2 both accumulate 2.5 units of variance, and covary proportional to their shared branch length of 1 unit of covariance.

IV. Brownian Motion on a phylogeny defines a *Multivariate Normal Distribution*, which is completely defined by a mean (a vector) and a variance (a matrix). The mean vector is simply the root state for every tip, while the variance is σ^2 **C.** We then estimate σ^2 and z_0.

V. Multivariate Brownian Motion is a very simple extension of Brownian Motion. The rate of evolution (σ^2) becomes a matrix (R) of form:

$$\mathbf{R} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

$\sigma_x{}^2$ represents the rate of evolution of trait x. $\sigma_y{}^2$ represents the rate of evolution of trait y, and $\sigma_{xy}{}^2$ represents the covariance of changes in x and y. If positive, they tend to change together. If negative, they change together, but in opposite directions. And if 0, then the two traits evolve independently. Furthermore, the mean/root state of the mvBM process becomes a vector, one mean for each trait.

10/17 PICs and PGLS

I. "The Comparative Method" - Attempting to find relationships between traits from comparing those traits among different species.
    A. For example, sister group comparisons: Compare pairs of sister species that differ between two traits. Does not require phylogeny beyond the identification of sister group pairs.

II. Phylogenetic Independent Contrasts (Introduced by Felsenstein 1985)
    A. Improves on "sister group" comparisons as it uses all the information in the data.
    B. Comes at a cost: Requires assuming a model for evolutionary history. Felsenstein proposed Brownian Motion.
    C. PICs are an algorithm for turning N tips into N-1 "independent contrasts". Assuming BM is the true model, these contrasts can be used in standard statistical analyses.
    D. Contrasts are simply the differences illustrated below standardized by their expected change under Brownian Motion (which is calculated from the rate of

evolution and the branch length).
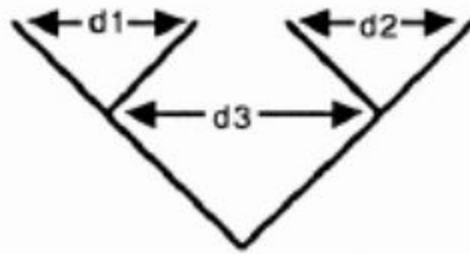
E. Once contrasts are calculated, simply perform a linear regression of PIC_Y ~ PIC_X with the intercept fixed at 0.

III. Phylogenetic Generalized Least Squares
   A. A more flexible modeling approach that ends up being equivalent to PICs under BM (which is sort of best thought of as a statistical coincidence, since the models are fairly different conceptually).
   B. Unlike PICs, does not assume trait X evolves by BM, only the residual variation of Y ~ X. In other words, X affects Y, but other unobserved, phylogenetically correlated traits also affect Y. Thus, we must account for these effects when estimating the effect of X on Y.
   C. PGLS is more flexible, and we can include Non-Brownian evolution via tree transformations:
      1. Lambda models: Scale the tree from 0 to 1, with 0 meaning no tree structure (left), to 1 meaning standard PGLS with BM (right). This is accomplished by multiplying the off diagonal elements of the tree matrix by lambda.

$$Y = \beta X + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

$$\varepsilon \sim N(0, \sigma^2 V)$$
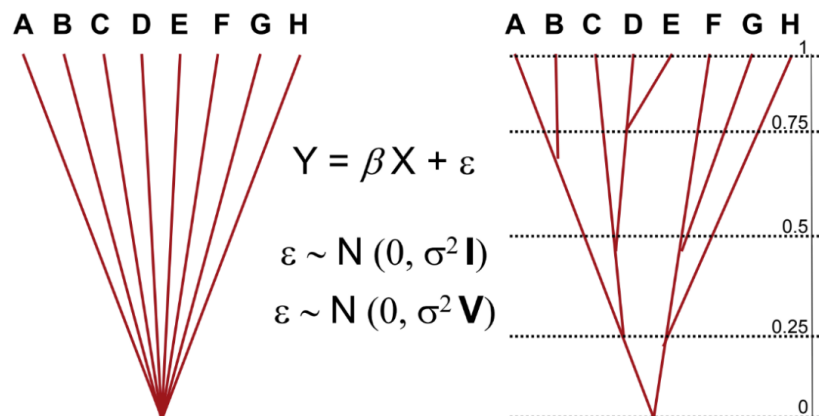
      2. Kappa models; Kappa = 1, standard PGLS with BM. Kappa = 0 transforms all branch lengths to length 1, which means that change is only proportional to the number of *speciation events*. In other words, change only happens at speciation. (But what if not all speciation events are represented in the phylogeny?)

3. Delta models: Node heights are raised or lowered to the power of delta. Delta = 1 is standard PGLS with BM. Delta > 1 means accelerating evolution. Delta < 1 means decelerating evolution.
4. We can also replace the BM part of standard PGLS with an OU model, which for ultrametric trees behaves much like PGLS + BM w/Lambda.

10/22 Modeling Adaptation with OU models
I. Ornstein-Uhlenbeck models are extensions of BM that include "pull" to an optimal state.
   A. Know the equation for an OU model and all its parameters
      1. Adds alpha and at least one theta to the root state.
   B. Know the equations for phylogenetic half-life and stationary variance and what they mean, and their units.
      1. Phylogenetic half-life = $Ln(2)/alpha$; amount of time it takes to get halfway to the optimum (theta)
      2. Stationary variance = $sigma^2/(2*alpha)$; the stationary distribution if the process is allowed to evolve for an infinite amount of time to a particular phenotypic optimum
   C. The Ornstein Uhlenbeck equation can be divided two components, a deterministic component (adaptation) and a stochastic component (random, Brownian evolution).
      1. If the half-life is very high (alpha very low) the model converges to Brownian Motion. The deterministic component goes away, and the optimum does not influence trait evolution, and *phylogenetic signal is high*.
      2. However, if the half-life is very low (alpha is very high), pull towards the optimum *erases phylogenetic covariance among closely related species*. If it is strong enough, the model becomes *White-Noise*, meaning there is no phylogenetic signal at all and every species is simply drawn randomly from the stationary distribution.
      3. Intermediate values of alpha that can be estimated from phylogenetic comparative data are typically between the height of the tree, and the youngest split on the phylogeny. Anything higher or lower than these limits should be interpreted as largely BM-like or WN-like (respectively).
      4. If fit to a single optimum OU model, half-life gives an overall estimate of phylogenetic signal.
   D. Emerges out of models of Genetic drift and selection to stationary peak (Lande), but probably not what we are modeling in macroevolution
II. Early Burst model - Decreasing rate of sigma^2 with time, but otherwise Brownian. Used to test for adaptive radiations.
   A. AC/DC model: more general version of early burst model that allows increasing or decreasing rates.

B. OU models often favored over BM or EB in empirical data. One reason is measurement error favors OU if unaccounted for.
C. Understand this figure:
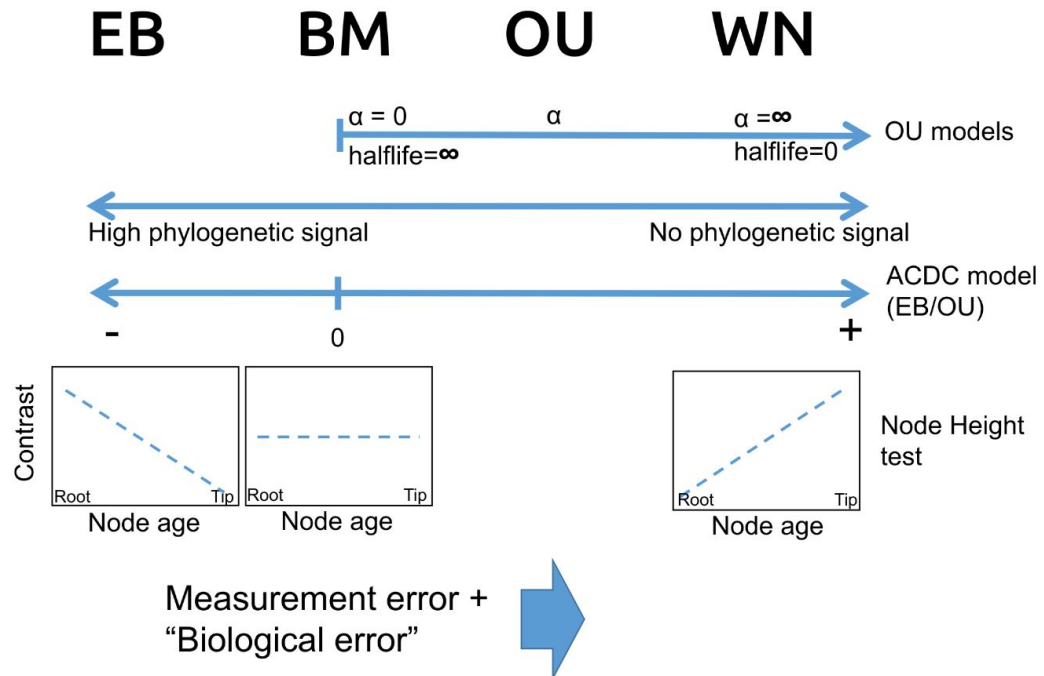
10/26 Testing adaptive hypotheses

I. Multi-optima OU models model shifts in phenotypic optima, generally based on reconstruction of one or a few predictors using ancestral state reconstructions (Butler and King 2004).

D. These optima or "regimes" shift along the tree. You can use parsimony reconstructions, or incorporate ancestral state uncertainty using *stochastic character mapping*.

E. Test optima shifts based on predictor (Butler and King 2004) or expanded to test shifts in optima, sigma^2 and/or alpha (Beaulieu et al. 2012).

F. Test different "regime paintings", estimating alpha, sigma^2 and theta, and compare these regime painting hypotheses using AIC, for example.

10/28 & 10/31 Finding adaptive shifts; Data-driven vs. Hypothesis testing frameworks

I. Understand the difference between data-driven and hypothesis testing frameworks, strengths and weaknesses. How can you be lead astray with model selection from a small set of a priori hypotheses? How can be lead astray by searching among all possible hypotheses?

II. Be able to explain how Hidden-state, reversible-jump etc. account for background heterogeneity unrelated to a particular hypothesis

III. Be able to explain the problem of Maddison and FitzJohn 2012 and Rabosky and Goldberg 2014.
    A. Maddison and FitzJohn: Darwin's scenario and character correlations
    B. Rabosky and Goldberg: Model selection when the correct model is not included:
        1. Caetano and Beaulieu's "wolf" and "no-wolf" example.
IV. Uyeda et al. 2018
    A. Darwin's scenario affects all phylogenetic comparative methods, would be good evidence of association if not for ascertainment bias (what is ascertainment bias?).
    B. Combining hypotheses tests with phylogenetic natural history one solution
    C. Simpson's paradox in comparative methods - When should you use PGLS vs. OLS?
    D. What's the difference between SLOUCH and PGLS?