Phylogeny-Based Analyses of Evolution ... with a Paleo-Focus
Part 1. Macroevolution of Traits (Feb 29, 2012)
David W. Bapst

Recommended Reading:

Models for Continuous Trait Evolution
“Essentially, all models are wrong, but some are useful.”
- George E.P. Box

Consider some process you are interested in: What do you think is the true model? Does any science or statistical analysis distinguish the true model? If so, what is the best way of doing that? If not, what is science about if not finding the true model? We should be mindful of these issues as scientists.

The models listed here describe how changes in traits are distributed across branches and taxa: for now, let us we are interested in looking at patterns of evolution where evolution is a function of time. Our trees will have branch lengths relative to divergence times.

Incorrect trees and branches are a big worry but, these analyses are almost completely insensitive to taxon sampling issues. Larger samples improve the parameter estimation. Trait observations from the fossil record can be readily included to better constrain trait values earlier in the tree.

Brownian Motion: The Basic Model (Felsenstein, 1985; 1988)
- 2 Parameters: Root trait and rate of change (variance of change)
- For example, to stochastically simulate a trait at time=x, where trait0 is the trait value at time=0:
  o trait[x] <- trait0 + rnorm(1, sd=sqrt(rate*x))
- For log-likelihood, can use multivariate normal distribution:
From O’Meara et al. (2006)

The sigma ($\Sigma$) in above is the phylogenetic variance-covariance matrix:

$$
\log(L) = \log \left[ \exp \left\{ -\frac{1}{2} [X - E(X)]' \Sigma^{-1} [X - E(X)] \right\} \right] \sqrt{\frac{(2\pi)^N \times \det(\Sigma)}} \tag{3}
$$

From O’Meara et al. (2006)

Because most trait evolution models can be described as branch-length transformations of a given phylogeny (or transformations of the vcv matrix), can use the above to find the likelihood under a wide range of models

**Pagel’s Delta and Time-Variant Rate Models** (Pagel, 1997; 1999)

- 3 Parameters: BM’s two parameters and delta
- Transformation of branch length / vcv matrix
  - Node heights raised to the power of Delta
- Basically fits a model where rates exponentially increase or decrease over time
- Similar model: ‘Early Burst’ (Harmon et al., 2010)
  - Fit linear or exponential rate change in time

**Pagel’s Lambda and Phylogenetic Signal Models** (Pagel, 1997; 1999)

- 3 Parameters: BM’s two parameters and lambda
- Transformation of vcv matrix diagonals
  - Off-diagonals (covariances) multiplied by lambda
  - lambda=1, no transformation, equal to Brownian Motion
  - lambda<1, traits have less covariance than expected by BM
  - lambda=0, co-variances are zero, equal to White Noise
  - lambda can be over 1, but covariances in the vcv matrix cannot exceed variances (so this will be true on trees with long terminal branches)
- Blomberg et al. (2003) presented $K$, an alternative that can exceed 1 to imply higher levels of signal than expected under Brownian Motion
K is calculated as the ratio of the trait mean square error (MSE) to the phylogenetic MSE, rescaled such that the MSE ratio expected under BM is equal to 1.

Losos (2008) argued that only reported values of K>1 should be termed “niche conservatism.”

- **Revell et al. (2008): what is phylogenetic signal?**
  - How much is the observed trait covariances explained by the covariances of the vcv matrix?
  - High signal: BM, Early Burst with early high rates
  - Low signal: Evolution with an optimum, constrained/bounded evolution, evolution with low interspecific heritability
  - High signal is sometimes **misinterpreted** as ‘high phylogenetic constraint’ or that trait variance is ‘explained’ by phylogeny

- Important: measurement error will look like decreased signal
  - This also means that models which are expected to produce low-signal trait values may be erroneously supported in noisy data

- Often used to assess phylogenetic patterns of extinction risk in studies of conservation (Purvis, 2008)

- Roy et al. (2009) found high signal in bivalve extinction rates

---

**Branch-length transformations on the Geospiza tree in geiger.**

**White Noise: The Under-Parameterized Model** (Smith et al., 2011)
- Two Parameters: Mean and Variance (just like normal dist.)
- As Brownian-Motion but ignores phylogenetic vcv matrix
• Trait value at time=x independent of value at time=0  
  o trait[x] <- rnorm(1,mean=mean,sd=sqrt(var))
• Assumes that lineages are evolving completely independently  
  o Equivalent to Pagel’s lambda=0 or OU1 with alpha=infinity
• No signal, no heritability, no similarity due to relatedness  
  o Example: Smith et al. (2011) find it supported in a cryptic species complex; interpret as very strong adaptation

**The Free Model: The Over-Parameterized Model** (Mooers et al., 1999)
• Like Brownian Motion where every branch has a different rate  
  o Number of parameters will be number of branches +1
• Little used since 1998, but perhaps useful to test if another set of models you are testing include or contain models complex enough to be sufficient descriptions of evolution in the study group  
  o i.e. if the Free Model fits best under AIC, compared to BM and EB, that (may) suggest a more complex process is going on

**Trend: Brownian Motion with a Slant**
• 3 Parameters: Root trait, rate of change, mean change  
• Equivalent to BM when the mean change = 0
• For example, to stochastically simulate a trait at time=x  
  o trait[x] <- trait0+rnorm(1,mean=meanChange,sd=sqrt(rate*x))

**Ornstein-Uhlenbeck 1: Evolution with a Single Attractor**
• Felsenstein, 1988; Hansen, 1997; Butler and King, 2004
• Multiple variants, most commonly fit is OU1 (single attractor)
• BM with two additional parameters: an attractor/optimal trait value (theta) and a strength of the pull to this optimum (alpha)
• These optima are generally assumed to be static and non-changing
• In most parameterizations, the root trait is dropped as a nuisance parameter and set equal to theta (but see OUwie)
• For example, to stochastically simulate a trait at time=x:  
  o trait[x]<-trait0+alpha*(theta-trait0)+rnorm(1,sd=sqrt(rate*th0))
• When alpha=0, becomes BM; when alpha=inf, becomes functionally identical to White Noise / no phylogenetic signal
• Optima often referred to as clade-wide adaptive regimes, zones, etc  
  o A dataset that is best fit by OU1 could probably also be interpreted as bounded evolution
• Expectation that more evolution will be apparent on later branches, as the pull to the optimum will erase phylogenetic signal resulting from evolution on earlier branches. Looks similar to low signal...  
  o Branches will overlap in traitspace (see below)
• Sidlauskas’ (2008) lineage density patterns suggest an OU1 / low-signal pattern of trait evolution

• Can estimate rate of approach to optimum using Hansen’s ‘phylogenetic half life’ method (1998)

• OU1 will often have better AIC for models with noisy trait data
  o Trait data often noisy in paleobiology: should we be worried?

<table>
<thead>
<tr>
<th>Brownian Motion</th>
<th>Early Burst</th>
<th>Trend</th>
<th>Ornstein-Uhlenbeck1</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Brownian Motion" /></td>
<td><img src="image2.png" alt="Early Burst" /></td>
<td><img src="image3.png" alt="Trend" /></td>
<td><img src="image4.png" alt="Ornstein-Uhlenbeck1" /></td>
</tr>
</tbody>
</table>

Traitgrams (Ackerly, 2009) are another handy way of visualizing the differences among various models. (Handmade, not simulated)

**Many Models are BM and OU with Parameters Varied**

• Different BM rates in different lineages (O’Meara et al., 2006)

  One way to think of this:
  A branch-length transformation where lengths are multiplied by different coefficients. This figure is of a tree where one subclade has x4 greater branch lengths. This would be equivalent to a x4 increase in the rate.

• BM on multiple traits, with correlated changes in rate of trait change (Revell and Collar, 2009)

• BM with Diversity-Dependent Rate (Mahler et al., 2010)

• Ornstein-Uhlenbeck with multiple optima (OU2, OU3, etc), alpha values
  o Butler and King (2004), Beaulieu et al. in press (“OUwie”)
  o Find best-fit combo of multi-optima (Hipp, 2007; mattice)

• Models with varying rate, optima and alpha: see R package OUwie and Bealieau et al. (in press)

• See Collar et al. (2009) for an example of comparing multi-rate models to multi-optima models using AIC
• Ornstein-Uhlenbeck where each species follows an optimum moving under Brownian motion (SLOUCH, Hansen et al., 2008)

**Price’s niche-filling model (Price, 1997)**

Price’s 1997 model, which is completely unrelated to Brownian Motion, instead having niches dropped randomly into a bivariate elliptical niche-space, with closest neighbor as ancestor. Species are static once placed. Obviously, this model has weird inheritance, partly due to multidimensionality.

• Freckleton and Harvey (2006) present a node-height test which can distinguish this pattern from BM (but what about from Early Burst?)

**Models on the Rise**

- Can simulate many models but have no likelihood function for them.
- Bounded BM – absorbing boundaries, reflective boundaries
  - Ackerly, 2009 simulated ‘bounded’ BM model, acted similarly as OU1
- Evolution with a Detractor (Alroy, 1998)
  - Ornstein-Uhlenbeck with a negative alpha?

**Comparative methods for Comparing Traits**

- Why do we often call phylogeny-based analyses of evolution “comparative methods”?

  - Comparative Methods (pre-1985) referred to a common type of analysis where phenotypic, physiological, habitats and life history variables were measured across a range of species and compared, via linear regression, to test for adaptation.
  - Felsenstein (1985) realized that species are not statistically independent; historical contingency needed to be accounted for
    - Worse-case: single early divergence defines relationship
    - Suggested phylogenetic independent contrasts
  - (Assuming BM) calculate differences between nodes on a tree, scaled to the branch lengths to produce n-1 statistically independent standardized contrasts
  - PICs can be compared in a regression, although should force regression through the origin (Garland et al., 1992)
Phylogenetic General Least Squares (Grafen, 1989)
- Use general least squares with a correlation matrix constructed under some model of trait evolution (more flexible than PIC)
  - Under BM, the vcv matrix is the correlation matrix
  - Can also use OU, delta, lambda, etc (Martins and Hansen, 1997)
- Garland et al. (1999, 2000), Rohlf (2001) ‘showed’ equivalency of PGLS under BM and PIC when regressions are forced through the origin; Blomberg et al. in press give the proof!
- Freckleton et al. (2002) and Revell (2010) recommend including lambda, so as to maximize estimation of possible covariances

**But** when should we account for phylogenetic non-independence?
- For example: is body size a predictor of poaching intensity? Is interspecific growth rates correlated with habitat degeneration?

http://www.mail-archive.com/r-sig-phylo@r-project.org/msg01594.html

Friedman, 2009: Extinction Intensity versus Morphological Traits
- Different sets of traits predicted extinction risk during K/T using non-phylogenetic and phylogenetic methods.
  - Treats extinction a proxy for another set of traits which predict extinction which does not include the traits he is actually studying (such as geographic range)
- What does this imply how these traits change on his tree?
- What does this mean for studies of extinction selectivity?

Revell (2010) has developed size-correction and PCA methods which take phylogenetic covariance into account

Price (1997) criticized these methods and presented a niche-filling model under which PICs obscured apparent relationship
- Although not presented as such, shows importance of model choice

**Ancestral Trait Reconstruction: Something You Probably Shouldn’t Do**
- Very basic ancestral trait reconstruction implicit in PICs
  - Any reconstruction method assumes some model, generally BM
- Widely used for some years until Schluter et al., 1997, presented a ML method which allowed calculation of confidence intervals
Unsurprisingly, these methods have been used less since because of the enormous uncertainty in their estimation

- Used to make ‘phylomorphospaces’ and ‘traitgrams’ (Sidlauskas, 2008; Ackerly, 2009) but shouldn’t make any strong inferences based on these plots given uncertainty (more aesthetic then science)
- Hunt (2007) tested evolution along lines of least resistance using reconstructions (didn’t use ML methods so no uncertainty measured)
- Finarelli and Flynn (2009) tested reconstructions of carnivore body size using fossil data and found poor fidelity of estimates
- Alroy (1998, 2000) and others use A-D change analyses, which bootstrap potential Anc-Desc pairs of fossil species. Alroy (2000) suggested using reconstructed values instead, but I would not recommend this given their uncertainty.

Estimating Rates and Comparing Models of Trait Evolution

- Standardized contrasts are unbiased estimate of rate (Garland, 1992)
- Pagel (1997;1999) presented methods for getting estimates of rates, and transformation parameters (lambda, delta, kappa)
- Mooers et al. (1999) compared trait evolution models via likelihood
- Butler and King (2004) developed method for fitting multi-optima models of Ornstein-Uhlenbeck (OUCH)
- Harmon et al. (2003) present disparity-through-time plots; more apt to refer to them as disparity-partitioning-through-time plots. They are basically another way of showing patterns better tested with model-fitting using AIC. Not clear what they mean for paleo-trees.
- O’Meara et al. (2006) developed ML framework using the multivariate normal distribution to test for rate differences (BROWNIE)
- Harmon et al. (2008) releases geiger which includes fitContinuous, which fits most other models
Where are we now? What have we learned about trait evolution?

- Two big scale studies of comparative datasets:
  - Most traits have signal but behavioral traits have less phylogenetic signal on average (Blomberg et al., 2003)
  - OU1 often best supported model, Early Burst supported very rarely in comparative data (Harmon et al., 2010)

- Some analyses found best-fitting OU models where taxa are far from estimated optimum, inching toward them (e.g. Collar et al., 2009)
  - Estes and Arnold (2009): are species on their optimum?

- Boettiger et al. (in press) found that model-fitting analyses using AIC can have low test power to distinguish models dependent on tree
  - Developed neat parametric boot-strapping technique to estimate test power using likelihood ratios

- In the fossil record, we probably don’t know our tree (including branch lengths) perfectly; this might add additional uncertainty which would lower test-power.

New Approaches: Out Now or Out Soon

- Finding shifts in rates without an a priori hypothesis of shifts
  - Thomas and Freckleton (2012) developed traitMedusa, which fits changes in rate of trait evolution under BM using stepwise AIC
  - Three lab groups independently developed Bayesian methods for inferring shifts in rates of trait evolution
    - Why Bayesian? Monte Carlo Markov Chains allow us to develop posterior samples of where shifts might be
    - traitMedusa only lets us know what the best ML model is
  - Revell et al. (2012) have MCMC method for a single shift
  - Eastman et al. (in press) have an rjMCMC method
    - rj = 'reversible jump'; the markov chain can jump between models of lower and higher complexity to sample different numbers of jumps
    - Estimate support for multiple numbers of shifts
  - Venditti et al. (2011) developed an MCMC method which they applied to the mammal supertree
  - None of these methods can consider multiple trees

- Dissecting evolutionary trends from species selection
  - FitzJohn (2010) modified BISSE framework to allow for quantitative traits such as body size (QuaSEE); see below

- Rates of trait evolution when phylogenies are completely known
  - If we don’t know the phylogeny, likelihood solution for BM is no good because we don’t know the covariances among taxa
A solution when we don’t know the likelihood function is use Approximate Bayesian Computation (ABC): estimate support for parameter values by doing repeated simulations
- What proportion of runs satisfies some observed criteria?
- Sidlauskaus (2007) and Hoerner (2011) developed methods which used simulations to estimate probabilities of rate parameters given some unresolved fish clades and then counted proportion of simulations that match real data as a likelihood
- Slater et al. (in press) developed an ABC MCMC method (MECCA)
  - Estimates the birth-death model that best fits the data
  - Simulates a tree under those birth-death rates
  - Simulates trait evolution under BM on the tree
  - Accepts the run parameters if they are close enough to the chosen criteria (summary statistics)

Speciation Trait Change and Testing Punctuated Equilibrium
- Okay, what if traits change only or mostly at branching events?
- Some assume this and use ‘unit-length’ trees
  - Trees where branches are all length = 1
  - Assumes that evolutionary change is directly correlated to the number of observed branching events on the tree
  - Even in the fossil record, we don’t observe all branching events
- Used in some analyses, such as PGLS and model fitting
- Agrawal and Fishbein (2008) and Moen (2006) tested for trends associated with speciation events or with molecular change
- Rather than assuming, can we measure proportion of change occurring at speciation rather than along branches?
- Ricklefs (2004): correlations between time, richness and variance
  - Ricklefs (2006) found otherwise (excellent simulation study!)
- Pagel’s Kappa: Measuring correlation with observed branching events
  - Model based on Brownian motion like lambda, so three parameters
    - root trait, rate of evolution and kappa
  - Another transformation metric, this one raises branches to the power of kappa, which can range from 0 to 1
  - So, at 1, all branches as observed; at 0, all branch lengths = 1
  - Still no consideration for unobserved branching events
- Bokma (2008) developed an ABC analysis which (like MECCA) fits birth-death models in order to estimate unobserved speciation events
  - Allow for separate estimation of components of trait evolution associated with time and speciation events
  - But assumes very simple diversification / trait evolution models
  - Cannot be directly applied to paleo-trees
- Gene Hunt (unpub) has fit models to fossil stratophenetic diagrams
  - Both Bokma and Hunt find support for both branching-event associated and time-gradual evolutionary change
Discrete Trait Evolution
- Testing for correlated evolution in discrete traits (Pagel, 1994)
- Basic markov models of character change (Lewis, 2001)
  - Variants with asymmetric or multistate traits
  - Many of the transformation models above can also be applied
- Rate estimation and model fitting with similar methods as above
  - Stochastic mapping allows for Bayesian approach to uncertainty in reconstruction ancestral state transitions (Huelsenbeck et al., 2003)
  - Phylogenetic signal of binary traits (Fritz and Purvis, 2011)
- Markov simulations useful for testing convergence, reversibility
  - Wagner and Erwin, 2006; Goldberg and Igic, 2008; Syme and Oakley, 2012
- Tangled asymmetries: asymmetrical trait change and traits which change diversification rates produce similar patterns
  - BISSE: Binary State Speciation and Extinction (Maddison et al., 2007) and other related methods evaluate trait evolution and diversification models simultaneously to deal with this
  - Rabosky and McCune (2008): A new age of species selection studies
- This is a very active area and I do not mean to short-change it here. However, all of the BISSE methods cannot be applied to the non-ultrametric trees used by paleontologists, so (for the moment) there is little need to discuss this any more

Making Use of Correlated Evolution: Inferring Traits
- Unobserved traits can be inferred based on an estimated correlated evolutionary relationship with another set of proxy traits you can observe (possibly even that one is incomplete across the tree)
  - i.e. trait A more likely to evolve given trait B
- Useful in fossil record for inferring things we can observe ever, like physiology, or only very rarely such as diet, but have possible proxies for in other traits: Organ et al., 2009; Zanno and Mackovicky, 2011; Schmitz and Motani, 2011

A short note on Biogeography
- Biogeography is generally treated as a trait
- Most geographic range evolution comparative analyses are based on the markov model
- Rabosky and Glor (2010) used BISSE to test models of diversification on islands, using island habitats as a trait

Why All of This is Relevant Even if You Never Touch a Tree
- Understanding these models useful even without phylogenies
Models are sets of assumptions which we can use to generate our worldview: Would this new method work with that model? Does this result agree with this commonly assumed model?

If you understand them well, can make predictions for making new methods that use those models. For example, estimate proxies for rates of trait evolution even without a tree being available.

**Analyses of Trait Evolution Using Paleo-Trees I Didn’t Mention**


**References for All of the Above:**


