"PRINCIPLES OF PHYLOGENETICS: ECOLOGY AND EVOLUTION"

Integrative Biology 200B Spring 2011

University of California, Berkeley B.D. Mishler

Feb. 3, 2011. **Quantitative character evolution within a cladogram (intro; ancestral trait reconstruction; phylogenetic conservatism)**

I. Continuous characters – ancestral states

Many traits of interest are measured on **continuous or metric scales** – size and shape, physiological rates, etc. Continuous traits are often useful for species identification and taxonomic descriptions; historically, they were also used in phylogenetic analysis through the use of clustering algorithms that can group taxa based on multivariate phenetic similarity. With the advent of cladistics, phylogenetic reconstruction per se has shifted entirely to discrete trait analysis, but continuous traits remain important in post-tree analysis.

In principal, ordinal traits take on discrete states while continuous traits are real numbers. In reality, there is a continuum from ordered discrete traits with only a few states (1,2,3,4,5 petals) to those with enough states that we may treat them as continuous (dozens to hundreds). It is also possible to 'discretize' traits into states if breaks are observed in the distributions for different species.

Distributions of continuous traits may take on any arbitrary shape, but certain distributions occur repeatedly, possibly reflecting common underlying 'natural' processes.

Normal distribution: sum of many small additive effects
Exponential distribution: product of many small multiplicative effects
Poisson distribution: frequencies of rare events in discrete intervals
etc.

This becomes important when we consider whether ancestral reconstruction of continuous traits should reflect an underlying evolutionary model of the process that describes or dictates trait evolution. Traits may be transformed to better meet an appropriate distribution (e.g. log-transform).

I.B. Parsimony methods for ancestral states

1. Linear or Wagner Parsimony: minimize the sum of absolute or linear changes along each branch (analogous to normal parsimony for discrete traits). The ancestral value at each node will be the *median* of the three values around it (two child nodes, one parent node). The root is a special case, where it will be the median of the two child nodes, as there is no parent node.

2. Squared-Change Parsimony: minimize the sum of squared changes along each branch. The ancestral value at each node will be the *mean* of the three values around it (two child nodes, one parent node). Weighted SCP can also be calculated, where the change along each branch is divided by its branch length before summing – a given change on a long branch is penalized less.

I.C. Brownian motion and maximum likelihood ancestral states

Brownian motion (BM) is the term for a random walk in a continuous valued variable. If a trait was determined by multiple, independent additive factors of small effect, and if each factor was mutating or changing at random (e.g., by drift), then the character change would constitute BM. Brownian motion is the starting point for discussions of continuous character evolution, for its simplicity and its close ties to parametric statistics based on normal distributions.

In Brownian motion the size of each step is drawn from a normal distribution with mean = 0 (no trend) and variance s^2 (= standard deviation s), where each step is a unit of time. When we consider Brownian motion as a process, this variance is viewed as a rate parameter, β. One of the fundamental principles of probability theory is that the variance of the sum of two random processes is the sum of their variances. In other words, if the variance of a brownian motion process is β after one time step, it will be β + β = 2β after two time steps. So the variance increases linearly with time.

If you apply BM to a large number of independent random walks, with time = t along each walk, then you can probably see that the variance of the resulting values at the tips of the walks will be tβ. What is less intuitive for most of us (if you are not used to statistical thinking) is that a single value resulting from a random walk also has a variance that refers to the underlying (and unobserved) distribution from which that value has been drawn.

Whether you know it or not, we all solve a maximum likelihood (ML) problem on a daily basis when we calculate the mean for a set of numbers. The mean of X (a set of numbers) is the sum of X divided by N, the number of values in X, right? Yes. Alternatively, the mean of X is the ML solution for the starting point of N random walks that end with values X. This can be solved from the following steps:

1) From the central limit theorem, we know that random walks generate values drawn from a normal distribution, with mean u and variance s^2.

2) The probability of each value of X, under a normal distribution is:

$$P(x) = \exp\left[-(x-u)^2/2s^2\right] \Big/ \; s\sqrt{2\pi}$$

3) The ML solution for u and s^2 are the values that maximize their cumulative probability over all values of x, and the cumulative probability is the product of the individual probabilities. A product of a series of values for P(x) looks pretty nasty, so instead let's take the sum of the log of P(x) (because the log of a product is the sum of the logs):

$$\log(P(x)) = \left[-(x-u)^2/2s^2\right] - \log\left(s\sqrt{2\pi}\right)$$

4) To maximize this, we can ignore the denominator ($2s^2$) and the second term, since they will be constants. And if we are maximizing the sum of negative terms, we can instead minimize the sum of the positive terms. So the mean of X is that value which minimizes:

$$\sum (x-u)^2$$

Look familiar!? It's the sum of squares of X. And now there's some magic, and the sum of squares is also how we calculate s^2, but we won't try to derive that as a ML problem here.

R script to solve for the mean of a set of numbers by finding the minimum of the sum of squares:

```
## enter a set of numbers in xx
xx = c(1,2,4,5)

## create a sequence of candidate values for the mean of xx
xu = seq(1,5,by=0.1)

## create a variable to hold the sum of squares
lxu = rep(NA,length(xu))

## loop through xu and calculate the likelihood score for each candidate value
as sum of squared deviations of xx from xu
for (i in 1:length(xu)) lxu[i] = sum((xx-xu[i])^2)

## plot the likelihood score vs. xu
plot(xu,lxu)

## find the minimum; print it out and compare it to the mean of xx
minxu = xu[which(lxu==min(lxu))]
print(minxu)print(mean(xx))
```

Ancestral states: Now we apply the same principles to solve for ancestral states under ML and Brownian motion, treating each ancestral state as the ML solution for a local BM process derived from that node, and finding the set of ancestral states that maximizes the likelihood over the entire tree. The branch lengths are key now, as the overall s2 value at each node is proportional to the BM rate parameter times the branch lengths.

Local likelihood solution: The ML reconstruction of ancestral states can be calculated as a local ML solution, based only on the trait data of tips descended from a node. This amounts to a recursive averaging process down the tree, except that at each node you calculate the weighted average of the two daughter nodes, weighted by the inverse of the square root of the branch length (more on that later)! (The everyday mean that we calculate for data sets is also a maximum likelihood solution for that value that minimizes the squared changes between the mean and the data points, i.e. minimizes the variance around the mean).

Global likelihood solution: The global ML solution uses information from the entire tree, including descendent taxa and all sister clades at and above (towards the root) a given node. Since ML is minimizing the sum of squared changes, the ancestral states found under the global likelihood solution are equivalent to the results under squared change parsimony. When a solution is found, a BM rate parameter is also calculated, based on the variance of the normal distribution per unit branch length (see Schluter al. 1997, top right of p. 1701). *The big difference between parsimony and ML is that ML techniques can provide confidence intervals on the ancestors.* Given the BM rate parameter, we can calculate a distribution of ancestral states (i.e. support limits) that are consistent with the observed data. The rather troubling result of much work in this area is that sometimes the error bars exceed the range of values observed in the terminal taxa (in other words the ancestor could be anywhere in the range of present-day trait

values, or even outside it!). See Schluter et al. 1997 Fig. 7 and Fig. 8 (error bars in Fig. 8 don't seem to show up in the pdf - check the printed journal).

I.D. Do ancestral state reconstructions work?

Three attempts to test methods: Oakley and Cunningham 2000, Polly 2001, Webster and Purvis 2002. The first used experimental bacteriophage lineages, directly examining properties of ancestral populations. Other two used comparisons with fossils. Polly found that fossil values were quite close to ancestral estimates, and closer than might be expected based on the confidence limits. Oakley and Webster papers both conclude that ancestral state methods perform very poorly—in both cases, there were significant evolutionary trends across the entire clade that caused the problems.

I.E. Citations on ancestral states for continuous characters

Cunningham CW, Omland K, Oakley TH. 1998. Reconstructing ancestral character states: a critical reappraisal. TREE 13:361-366

Losos J. 1999. Commentaries : Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. Animal Behaviour 58:1319-1324

Maddison W. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. Syst. Zool. 40:304-314

Martins EP. 1999. Estimation of ancestral states of continuous characters: a computer simulation study. Syst. Biol. 48:642-650

Oakley TH, Cunningham CW. 2000. Independent contrasts succeed where ancestral reconstruction fails in a known bacteriophage phylogeny. Evolution 54:397-405

Polly PD. 2001. Paleontology and the comparative method: ancestral node reconstructions versus observed node values. Amer. Nat. 157:596-609

Schluter D, Price T, Mooers A, Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699-1711

Schultz T, Cocroft R, Churchill G. 1996. The reconstruction of ancestral character states. Evolution 50:504-511

Swofford DL, Maddison WP. 1987. Reconstructing ancestral states under Wagner parsimony. Math. Biosci. 87:199-229

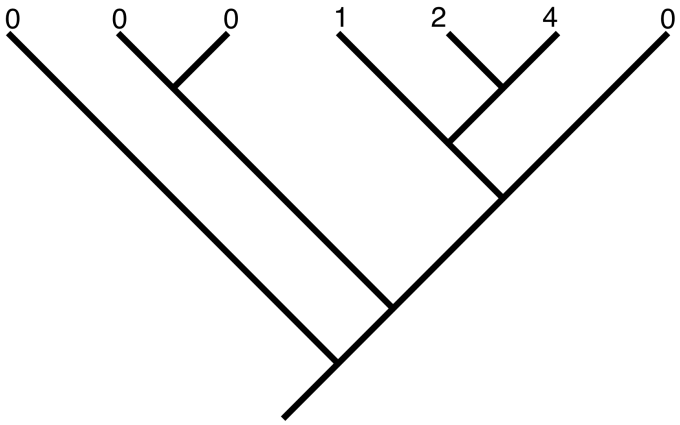
Webster AJ, Purvis A. 2002. Testing the accuracy of methods for reconstructing ancestral states of continuous characters. Proc. Roy. Soc. London Ser. B 269:143-149

R script to generate some random walks and their variance over time:

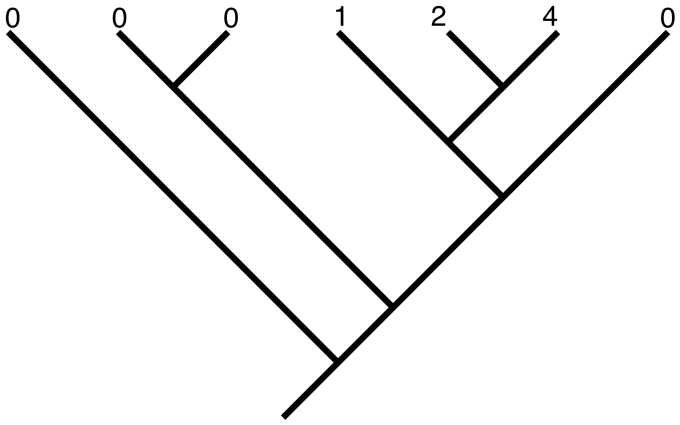
```
N = 10 #number of walks
L = 10000 #length of each walk
x = replicate(N,cumsum(rnorm(L)))
xr = range(x)
plot(x[,1],type='n',ylim=xr)
for (i in 1:10) for (j in 1:9999) lines(c(j,j+1),c(x[j,i],x[j+1,i]))


## plot variance over time of the outcome of the N walks
vx = apply(x,1,var)
plot(vx,xlab='time',ylab='variance of x',type='l')
```

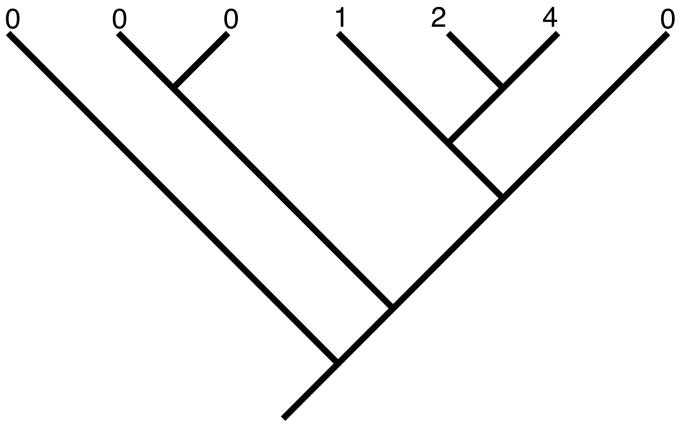
Linear Parsimony



Squared change parsimony (no BL, or equivalently all BL = 1)



ML under Brownian motion (with BL as drawn)




```
## 7 taxon example, ace solution
require(ape)
tree =
"(t1:5,((t2:1,t3:1):3,((t4:2,(t5:1
,t6:1):1):1,t7:3):1):1);"

phy = read.tree(text=tree)
plot(phy)
x = c(0,0,0,1,2,4,0)
ace(x,phy)
```

II. Phylogenetic signal

Another way to think about the fit to Brownian motion is the broader question of whether traits exhibit phylogenetic signal, and if so how the patterns of similarity among related species may reflect expectations under different evolutionary models.

II.A. Pattern and Process

Pattern is not process. No matter how many times we hear and repeat this phrase, the temptation to infer the action of specific processes from pattern alone is inevitable. In comparative biology, it has been all to common to equate stasis (pattern) with 'inertia' or 'constraint' (process?) and evolutionary change (pattern) with the action of selection (process). It is essential at the outset to see these dichotomies of pattern and process as orthogonal, setting up a 2x2 table:

	Pattern of trait evolution	
Evolutionary process	Stasis	Change
Adaptive	Stabilizing selection Fluctuating directional selection	Directional or disruptive selection
Non-adaptive	Lack of genetic variation (= constraint?) Antagonistic correlations among traits under selection Swamping by gene flow	Mutation Genetic drift Genetic correlations with traits under selection

In this case, I equate the term 'adaptive' with patterns resulting from the action of selection. It is not a statement about the functional significance of a particular trait, nor is it the same as stating that a particular trait represents an 'adaptation'.

It is critical to remember, as well, that these processes are neither exclusive nor independent of each other. For example, the action of drift and selection in small populations may simultaneously contribute to changes in gene frequency, and it may be difficult or impossible in an individual instance to separate their contributions. Replication in space (populations) or time (across generations) is critical, as selection is the only process we currently know of in evolutionary biology that can cause repeated, heritable change in the same direction in response to environmental conditions. Note also that selection leads to the loss of genetic variation. Thus, strong selection on a trait may lead to fixation of genes responsible for variation, such that future maintenance of the trait is due to the absence of sufficient variation (even if the selective context shifts).

As discussed in the Blomberg and Garland paper, there is a long history of usage of terms such as phylogenetic inertia or constraint, and other more recent terms such as phylogenetic effect and conservatism. To a greater or lesser degree, all of these terms imply the action of some underlying process. The term *phylogenetic signal* is useful as the implication is clear that we are only speaking about the pattern. I also like phylogenetic conservatism, though it is now taking on its own intellectual baggage and may suggest particular hypotheses about process (e.g. see Losos 2008 Ecology Letters).

II.B. Quantifying phylogenetic signal for continuous traits

Using the parsimony models discussed above in section I.B., we could look for phylogenetic signal via Monte Carlo methods similar to approaches discussed with qualitative characters. We could randomly permute the values for OTUs and see how our "real" tree length compares to the random values.

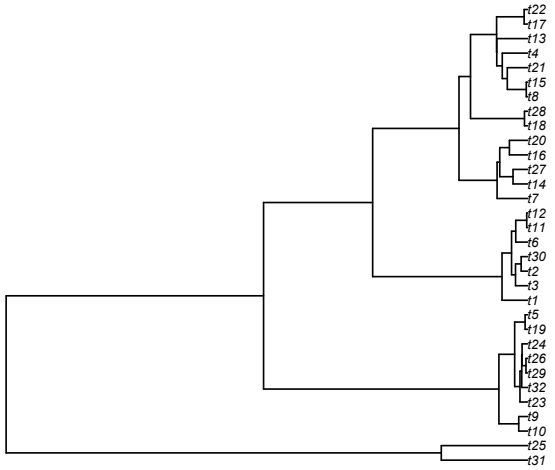
Several statistical have been developed to quantify phylogenetic signal, and test for significance relative to a null model of no signal. Freckleton et al. (2002) discuss the λ parameter, which is derived from Pagel's model fitting methods. Blomberg et al. (2003) proposed the K statistic. λ and K have the same general interpretation:

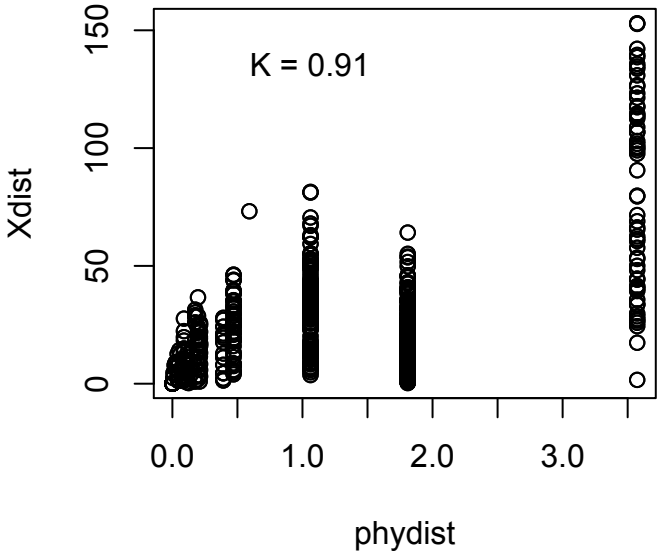
λ, $K < 1$ less signal than expected under Brownian motion
λ, $K = 1$ expectation under B.M.
λ, $K > 1$ more signal than expected under B.M.

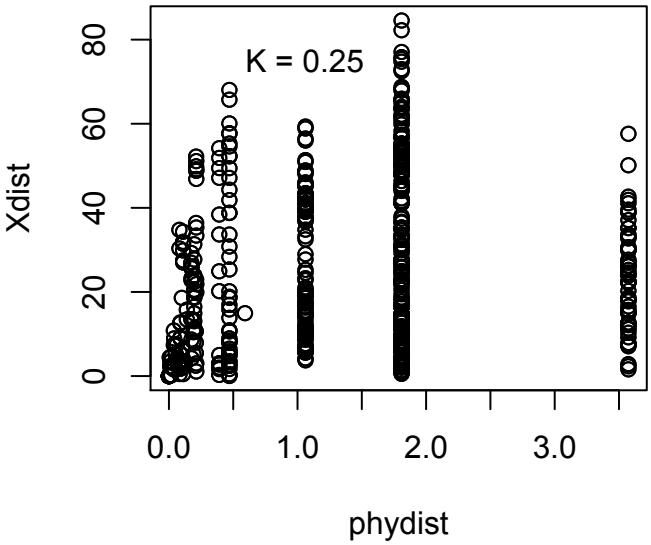
Quantitatively, one of the main differences is that λ has a maximum around 1.1 or 1.2, so it has little resolution to describe patterns with a high degree of signal (Freckleton et al. don't report values > 1). K on the other hand has a theoretical maximum of infinity and can provide some insight into differing degrees of signal greater than expected under Brownian motion.

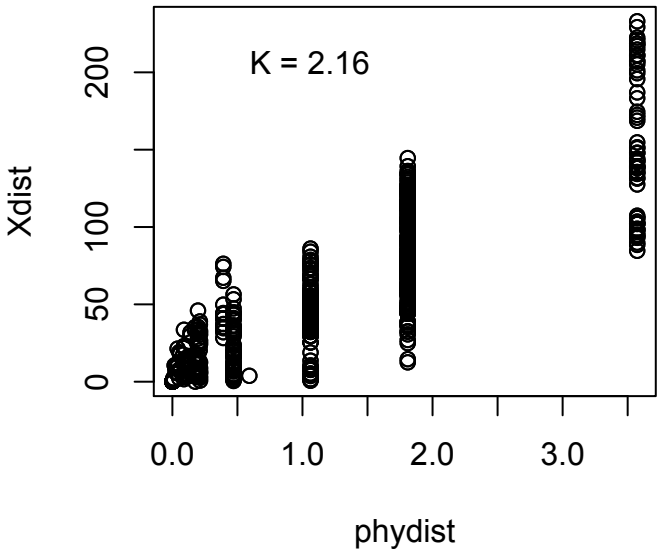
Fortunately, there is a simple graphical interpretation of these statistics, based on a plot of the difference in trait values between species pairs (all pairwise comparisons, not just sister taxa!) versus the phylogenetic distance between the taxa. Remember from Brownian motion, the expectation of the squared difference between trait values goes up linearly with the phylogenetic distance. It turns out that this translates approximately into a filled half triangle in a plot of the absolute difference vs. phylogenetic distance for all species pairs. Concave curvature of the upper surface of this plot means that close relatives are less similar than expected, compared to distant relatives ($K < 1$). Convex curvature, or an absence of points in the lower right, means that close relatives are more similar than expected ($K>1$). The details of the plot structure will depend on tree topology, but in general examining these plots is quite informative. Significance tests for the correlations in such plots must be tested using Mantel tests for matrix correlations (another topic!).

THE TREE


t22
t17
t13
t4
t21
t15
t8
t28
t18
t20
t16
t27
t14
t7
t12
t11
t6
t30
t2
t3
t1
t5
t19
t24
t26
t29
t32
t23
t9
t10
t25
t31


K = 0.91
Xdist
phydist
0
50
100
150
0.0
1.0
2.0
3.0


K = 0.25
Xdist
phydist
0
20
40
60
80
0.0
1.0
2.0
3.0


K = 2.16
Xdist
phydist
0
50
100
200
0.0
1.0
2.0
3.0

Another approach to asking about phylogenetic signal and patterns is to ask whether the trait data fit an alterantive model of trait evolution, and whether the parameters of this model then shed light on the processes influencing the trait and its evolutionary history. One of the very important alternative models is to introduce stabilizing selection, i.e. a model in which there is a 'pull' towards some central optimal value (which may fluctuate along different lineages), and Brownian motion reflects the random processes and/or the excursions in the action of selection around this underlying optimum. The stabilizing selection model is known as an Ornstein-Uhlenbeck stochastic process, so you will see references to OU models (and there is an R library called 'ouch' for OU models).

There are many evolutionists who believe that stabilizing selection is the overarching cause of evolutionary stasis and the maintenance of similarity among close relatives. Paradoxically, an OU process with a single optimum generates traits with K<1, so stabilizing selection reduces phylogenetic signal, which is not intuitive at first! Estes and Arnold (2007, American Naturalist) offer an important discussion of stabilizing selection and apply it to a large data set compiled by Gingerich (1983, Science) on rates of morphological evolution (mostly from fossils, not comparative data).

There are many other possible models, which have received more or less attention in the literature.

Mode	Model	Verbal explanation
Brownian	$x_{t+1} = x_t + N(0,s)$	random walk
Brownian + trend	$x_{t+1} = x_t + N(t,s), t \neq 0$	random walk with a trend
Bounded Brownian	$x_{t+1} = x_t + N(0,s)$ if x_{t+1} < minX or > maxX, resample	random wak with lower and/or upper limits to trait values
Proportional	$x_{t+1} = x_t * LN(\log mean = 0, \log sd = s)$	multiplicative random walk (equivalent to a random walk on log of trait)
ACDC	$x_{t+1} = x_t + N(0,s)$ $s = s_0 \gamma^{-t}$ $\gamma < 1$, accelerating $\gamma > 1$, decelerating	random walk with an accelerating or decelerating rate parameter
OU (Ornstein-Uhlenbeck)	$x_{t+1} = \mu + \theta(x_t - \mu) + N(0,s)$ μ = trait optimum $\theta = 1$: brownian $\theta = 0$: complete stabilizing selection	stabilizing selection; random walk with a 'pull' towards the optimum; theta is strength of the pull
Speciational	Brownian, with change only at time of speciation	random walk, but only one change per speciation event (i.e. branch length = 1 between each speciation event)
Punctuational	Brownian, with change only at time of speciation and only in one daughter lineage	like speciational, but change only occurs in one daughter lineage and the other exhibits stasis

Empirical reviews of phylogenetic signal:
Freckleton et al. 2002
Blomberg et al. 2003

Literature:

Abouheif, E. 1999. A method for testing the assumption of phylogenetic independence in comparative data. Evol. Ecol. Res. **1**:895-909.

Ackerly, D. D. 1999. Phylogeny and the comparative method in plant functional ecology. Pages 391-413 *in* M. Press, J. Scholes, and M. G. Barker, editors. Physiological plant ecology. Blackwell Scientific, Oxford, United Kingdom.

Ackerly, D. D., and M. J. Donoghue. 1995. Phylogeny and ecology reconsidered. J. Ecol. **83**:730-732.

Ackerly, D. D., and M. J. Donoghue. 1998. Leaf size, sapling allometry, and Corner's rules: a phylogenetic study of correlated evolution in maples (Acer). Amer. Nat. **152**:767-791.

Archie, J. 1989. Homoplasy excess ratios: New indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. SystZool. **38**:253-269.

Archie, J. 1996. Measures of homoplasy. Pages 153-188 *in* M. J. Sanderson and L. Hufford, editors. Homoplasy: the recurrence of similarity in evolution. Academic Press, San Diego.

Blomberg, S. P., and T. Garland, Jr. 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. J. Evol. Biol. 15:899-910.

Blomberg, S. P., T. Garland, Jr, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution **57**:717-745.

Burt, B. B. 2001. Evolutionary stasis, constraint and other terminology describing evolutionary patterns. Biol. J. Linn. Soc. **72**:509-517.

Coddington, J. 1994. The roles of homology and convergence in studies of adaptation. Pages 53-78 *in* P. Eggleton and R. Vane-Wright, editors. Phylogenetics and ecology. Academic Press, London.

de Queiroz, K. 1996. Including the characters of interest during tree reconstruction and the problems of circularity and bias in studies of character evolution. AmerNat. **148**:700-708.

Desdevises, Y., P. Legendre, L. Azouzi, and S. Morand. 2003. Quantifying phylogenetically structured environmental variation. Evolution **57**:2647-2652.

Diaz-Uriarte, R. & Garland, T. (1996) Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Systematic Biology,* **45,** 27-47.

Diniz-Filho, J. A. F., C. E. Ramos de Sant'Ana, and L. M. Bini. 1998. An eigenvector method for estimating phylogenetic inertia. Evolution **52**:1247-1262.

Estes S and Arnold SJ. 2007. Resolving the paradox of stasis: Models with stabilizing selection explain evolutionary divergence at all timescales. Amer. Nat. 169: 227-244.

Farris, J. 1989. The retention index and the rescaled consistency index. Cladistics **5**:417-419.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of the evidence. Amer. Nat. **160**:712-726.

Garland, T., Jr., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst. Biol. **41**:18-32.

Gittleman, J., and H.-K. Luh. 1992. On comparing comparative methods. Ann. Rev. Ecol. gittSyst. **23**:383-404.

Gittleman, J., and H.-K. Luh. 1994. Phylogeny, evolutionary models and comparative methods: a simulation study. Pages 103-122 *in* P. Eggleton and R. Vane-Wright, editors. Phylogenetics and ecology. Academic Press, London.

Gittleman, J., and M. Kot. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. Syst. Zool. **39**:227-241.

Givnish, T. 1997. Adaptive radiation and molecular systematics: issues and approaches. Pages 1-54 *in* T. J. Givnish and K. J. Sytsma, editors. Molecular evolution and adaptive radiation. Cambridge

UnivPress, New York.
Hansen, T. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341-1351.
Hansen, T. F., and S. H. Orzack. 2005. Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: The need for controlled comparisons. Evolution 59:2063-2072.
Harvey, P. H., and T. H. Clutton-Brock. 1985. Life history variation in primates. Evolution **39**:559-581.
Harvey, P., A. Read, and S. Nee. 1995. Further remarks on the role of phylogeny in comparative ecology. JEcol. **83**:733-734.
Harvey, P., A. Read, and S. Nee. 1995. Why ecologists need to be phylogenetically challenged. JEcol. **83**:000-000.
Kochmer, J., and S. Handel. 1986. Constraints and competition in the evolution of flowering phenology. EcolMonogr. **56**:303-325.
Leroi, A. M., M. R. Rose, and G. V. Lauder. 1994. What does the comparative method reveal about adaptation? Amer. Nat. **143**:381-402.
Lord, J., M. Westoby, and M. Leishman. 1995. Seed size and phylogeny in six temperate floras: constraints, niche conservatism, and adaptation. Amer. Nat. **146**:349-364.
Martins, E. 1996. Phylogenies, spatial autoregression, and the comparative method: a computer simulation test. Evolution **50**:1750-1765.
McKitrick, M. C. 1993. Phylogenetic constraint in evolutionary theory: Has it any explanatory power? Ann. Rev. Ecol. Syst. **24**:307-330.
Morales, E. 2000. Estimating phylogenetic inertia in Tithonia (Asteraceae): A comparative approach. Evolution **54**:475-484.
Orzack, S. H., and E. Sober. 2001. Adaptation, phylogenetic inertia, and the method of controlled comparisons. Pages 45-63 *in* S. H. Orzack and E. Sober, editors. Adaptationism and optimality. Cambridge University Press, Cambridge.
Peterson, A. T., J. Soberon, and V. Sanchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. Science **285**:1265-1267.
Rees, M. 1995. EC-PC comparative analyses. JEcol. **83**:891-892.
Ridley, M. 1992. Darwin sound on comparative method. TREE **7**:37.
Sanderson, M. 1991. In search of homoplastic tendencies: Statistical inference of topological patterns in homoplasy. Evolution **45**:351-358.
Starck, J. 1998. Non-independence of data in biological comparisons a critical appraisal of current concepts assumptions and solutions. Theory Biosci. **117**:109-138.
Wanntorp, H.-E. 1983. Historical constraints in adaptation theory: traits and non-traits. Oikos **41**:157-160.
Westoby, M., M. Leishman, and J. Lord. 1995. Further remarks on phylogenetic correction. JEcol. **83**:727-729.
Westoby, M., M. Leishman, and J. Lord. 1995. Issues of interpretation after relating comparative datasets to phylogeny. J. Ecol. **83**:892-893.
Westoby, M., M. Leishman, and J. Lord. 1995. On misinterpreting the "phylogenetic correction". JEcol. **83**:531-534.