The adequacy of trait models and the rise of angiosperm functional diversity

Matthew W. Pennell[1,2], William K. Cornwell[3,4], Luke J. Harmon[1]

[1] Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844, U.S.A.

[2] mwpennell@gmail.com

[3] VU University

[4] University of NSW

Abstract

Abstract here

There are known knowns; there are things we know we don't know.
There are known unknowns; that is to say, there are things that we
now know we don't know. But there are also unknown unknowns
— there are things we do not know we don't know.

—Fmr. U.S. Secretary of Defense Donald Rumsfeld

Introduction

Angiosperms are one of the most spectacular radiations in the history of the
earth. The >300,000 species which make up the clade have diversified into
a huge array of functional forms, from the grasses of the Serengeti to the
trees of the Amazonian rainforest, and have come to dominate most terrestrial
ecosystems. As researchers interested in macroevolution, we would like to
understand how such trait diversity came to be — the who, what, when, where
and whys of deep time. These questions need to be placed in a phylogenetic
context if we are to make substantiave progress towards addressing them. For
example, seed size, which is a proxy for a plant's life–history strategy, varies
X orders of magnitude across angiosperms (**?????**) [COMPLETE EXAMPLE
HERE]

The last few decades have seen a tremendous growth in statistical meth-
ods that make use of phylogenetic trees and comparative data to address
macroevolutionary questions (for a recent review see **?**). We can ask whether
traits are evolutionary correlated with one another (e.g. Felsenstein, 1985; **?**),
whether different lineages have evolved at different rates (e.g. O'Meara et al.,

2006; Eastman et al., 2011), what the predominant mode of evolution has been (e.g. Hansen and Martins, 1996; **?**; Harmon et al., 2010), how the mode has varied across clades (e.g. Butler and King, 2004; Beaulieu et al., 2012), and how trait evolution has influenced the diversification of lineages (e.g. **??**). However, all of these necessarily rely on our ability to model the evolution of traits along a phylogeny; and as eloquently articulated by Hunt (2012), any statements regarding the rate (tempo) of evolution are contigent about the particular model of change (mode).

A number of models of trait evolution have been proposed (see **??**). Brownian motion (BM), in which evolution proceeds via an undirected random walk such that the variance in the trait value accumulates proportional to time, is the oldest, and most commonly applied model of continuous character evolution. Originally developed as a model for estimating phylogenetic trees from allelic frequencies (Edwards and Cavalli-Sforza, 1964), BM has been used as a more general model of phylogenetic change in phylogenetics (Felsenstein, 1973; Thompson, 1975) and is the underlying model in Felsenstein's phylogenetically independent contrasts (PICs; Felsenstein, 1985) for investigating correlated evolution. Another model is the Ornstein–Uhlenbeck process (OU, or "Hansen" model; Hansen, 1997) which can be loosely described as "evolution on a spring" — variance accumulates at a rate σ^2, but is "pulled" towards a mean value θ with some strength α. Alternate models that have been considered include: the "Early Burst" (EB; Blomberg et al., 2003; Harmon et al., 2010; Slater and Pennell, in revision) model, in which most of the evolution occurs early in the clades history; a BM model with a trend in the mean trait value (Hunt, 2006); models depicting jumps in phenotypic space (**??**); and

other variations on the models described above (e.g. Pagel, 1997; **?**; Butler and King, 2004; O'Meara et al., 2006; Eastman et al., 2011; Beaulieu et al., 2012; Slater, in revision).

The general procedure for using a macroevolutionary models is to first compare amongst a set of candidate models, using likelihood ratio tests or information criteria. The preferred model is then used in one of two ways: 1) make inferences directly from the model fit by interpreting the observed pattern (e.g. Harmon et al., 2010; **?**); or 2) use the model to test other evolutionary hypotheses (e.g. **??**). In either case, there are a number of important questions that need to be considered. One is that of interpretation — what can our models tell us about the process of evolution (Hansen and Martins, 1996; **?**; **?**)? Another is statistical — is the model we are using capturing sufficient amount of the variation to address the question we are interested in? The latter question, known in statistical terminology as model adequacy, is the subject of our investigation here.

Consider again the example of seed size evolution in angiosperms. We may want to ask, for instance whether the variation in seed size we observe today accumulated early in the history of some clade of interest, which may suggest an adaptive radiation (**????**Slater and Pennell, in revision). To do we could fit mulitple models of trait evolution — an Early Burst model (Blomberg et al., 2003; Harmon et al., 2010) and a BM model. We could then compare the model fits using some model selection criterion (e.g. AIC; **?**). Once we have the best–fitting model in hand, we would like to be able to draw inferences from it as to the general tempo of seed size in our group. But before we can do

5

so, we would want to ensure that our model is actually capturing the relevant variation — that is, is our model adequate?

In many statistical applications, assessing the absolute fit of the model *a posteriori* is a routine procedure (**?**). Before drawing inferences about the parameters of the model, we want to know whether the model we used is adequately capturing the relevant patterns in the data. Model adequacy has been investigated for models of sequence evolution for the purposes of inferring phylogenetic trees (e.g. **?????????**) and forms the basis for the Decision–Theoretic approach to model selection in phylogenetics (**???**). However, in phylogenetic comparative methods, model adequacy has until very recently been largely neglected and its potential importance, generally underappreciated (but see below).

Boettiger et al. (2012) developed a "phylogenetic monte carlo" approach to assess when a given comparative data set contained enough information to distinguish between two candidate models. In brief, their procedure was as follows: 1) select 2 candidate models $\mathcal{M}_0$ and $\mathcal{M}_1$; 2) fit $\mathcal{M}_0$ and $\mathcal{M}_1$ to the data using Maximum likelihood; 3) use the MLE for model parameters $\hat{\Theta}$ to simulate n data sets; 4) for each of the n data sets simulated under $\mathcal{M}_0$, fit both $\mathcal{M}_0$ and $\mathcal{M}_1$ and calculate difference in likelihood values $\delta = -2(L_{\mathcal{M}_0} - L_{\mathcal{M}_1})$; and 5) compare distribution of δ for datasets simulated under $\mathcal{M}_0$ with distribution of δ for datasets simulated under $\mathcal{M}_1$. Boettiger et al. (2012) demonstrated that their approach had much better "classical" statistical properties (specifically Type–I and Type–II errors) compared to using Information Theoretic methods of model selection, such as AIC (**?**), AICc (**?**), and BIC (Schwarz,

1978). (Though we note that this is not an entirely fair comparison; Information Theoretic approaches differ philosophically from frequentist approaches to model selection [?] and do not really have Type–I and Type–II error rates in the same way likelihood ratio tests [?] are expected to have.)

Slater and Pennell (in revision) also focused on the case of comparing two–models. Their method differs from that of Boettiger et al. (2012) in that theirs was a fully Bayesian posterior predictive approach specifically aimed at detecting "early bursts" of trait evolution (*sensu* ??Harmon et al., 2010). They sampled from the joint posterior distribution of the parameters, simulated data under the sampled parameters and then used two alternative summary statistics — the relationship between the logarithm of the phylogenetic independent contrasts and the height above the root that the contrasts was inferred (a.k.a. the "node height test"; ?), and the Mean Disparity Index (MDI; ??) — to evaluate...

However, both of these cases used a simulation based approach to assess whether the data was informative enough to select amongs two candidate models. A broader question is whether a given model is a good fit to the data on its own terms — compared to the universe of possible models we could consider.

The aim of our paper is two–fold. First, we wanted to address a fundamental macroevolutionary question: what are the major patterns in the evolution of plant functional traits? We focused on five important ecological traits — seed mass, leaf size, maximum height, specific leaf area and leaf nitrogen content (the latter two being proxies for the "leaf economic spectrum", which

7

characterizes the plant's hydrodynamics; **??**) — which together encompass the major axes of life history and functional variation in vascular plants (**?**). Investigating this empirical question forced us to address more theoretical ones, relating to the second aim of this paper: do our evolutionary models capture meaningful patterns when applied at this scale and how are we to know?

In this paper, we develop a general approach to assessing the adequacy of trait evolutionary models for continuous characters.

Methods and data

Methodology

Our approach is based on the use of Felsenstein's 1973; 1985 Phylogenetic Independent Contrasts (PIC) method, which we will briefly review (for more details, see **??**). We have observed trait values $X_1, X_2, \ldots, X_N$ at the tips of a phylogenetic tree $\mathcal{T}$ consisting of N species. Due to shared history of ancestry between the tips, $X_1, X_2, \ldots, X_N$ are not independent observations. To deal with this problem, Felsenstein suggested taking $N-1$ contrasts $c_1, c_2, \ldots, c_{N-1}$, the differences $X_i - X_j$ between the observations at tips i and j. If we assume a BM model of trait evolution, in which variation accumulates directly proportional to time, these contrasts will be Independent and Identically Distributed (I.I.D.), hence the name PICs. The procedure can be described algorithmically. 1) Take the contrast $c = X_i - X_j$ at node k, where k is the most recent common ancestor of tips i and j. 2) Standardize the contrast by its dividing by its standard deviation, which under BM is $\sqrt{v_i + v_j}$, the square root of sum of

the branch lengths leading to *i* and *j*, 3) Estimate a trait value for the ancestral node *k* by taking the mean of its descendants' trait values, weighted by their branch lengths

$$X_k = \frac{(1/v_i)X_i + (1/v_j)X_j}{1/v_i + 1/v_j}.$$ (1)

4) Lengthen the branch below *k* by $v_i v_j / (v_i + v_j)$, in order to account for error in the estimation of *k*. Iterating across all nodes in the phylogeny, the result is a set of contrasts **c**, which, as stated above, will be I.I.D., *only if the true model which generated the observations was BM*. As our method described in this paper essentially evaluates whether this condition holds, we will refer to **c** as contrasts, rather than PICs throughout.

Summary statistics

To assess model adequacy we have chosen 6 summary statistics $\mathcal{S}_1, \ldots, \mathcal{S}_6$:

1. $\overline{\mathbf{c}^2}$: the mean of the squared contrasts. This is equivalent to the Restricted maximum likelihood (REML) estimate of the Brownian motion rate parameter σ^2 (**??**). We chose this statistic to capture variation in the rate of trait evolution.

2. D_c: the D–statistic obtained from Kolmolgorov-Smirnoff [SPELLING] test (**?**) from comparing the distribution of contrasts to that of a normal distribution with mean 0 and standard deviation $\sqrt{\overline{\mathbf{c}^2}}$. This is the expected distribution of the contrasts under BM (Felsenstein, 1985; **?**). We

9

chose this to capture deviations from normality, such as would be produced if traits evolved via a "jump diffusion" type model (**??**), in which trait evolution may occasionally occur at rates much greater than the background rates (see **?**).

3. $\text{var}(|\mathbf{c}|)$: the variance in the absolute value of the contrasts. This was chosen to capture heterogeneity in the rate of trait evolution.

4. m_{cv}: the slope resulting from fitting a linear model between the absolute value of the contrasts and their expected variances (**?**). Each contrast has an expected variance equal to XX (Felsenstein, 1985). Under a model of BM, we expect no relationship between these. In using this, we are asking whether the contrasts are larger or smaller than we expect based on their branch lengths. If, for example, more evolution occured per unit time on short branches than long branches, we would observe a negative slope.

5. m_{ca}: the slope resulting from fitting a linear model between the absolute value of the PICs and the inferred ancestral state (**?**). We estimated the ancestral state using the least–squared method suggested by (Felsenstein, 1985) as this uses an identical procedure to that done when estimating PICs. This statistic will allow us to evaluate whether there is variation in rates relative to the trait value (e.g. do larger organisms evolve faster?)

6. m_{ct}: the slope resulting from fitting a linear model between the absolute value of the contrasts and the height above the root at which they are

inferred. This is alternatively known as the node height test (?Slater and Pennell, in revision) for detecting early bursts of trait evolution and has been been previously used to assess the fit of BM models (?).

Rescaling the phylogeny

While the above summary statistics are appropriate for a BM model of trait evolution, the same will not be true of alternative models. That is because under alternative models, we no longer expect PICs to have I.I.D. properties. Our solution to the problem is to use the estimated parameters of a more complex model Θ to create what we term a "unit tree". A unit tree is defined as a tree in which if the model we fit is the generating model, the data at the tips will be distributed as it would be under a BM process with a rate σ^2 equal to 1. Below, we define this concept mor formally. Any phylogenetic tree $\mathcal{T}$ can be completely described by a $N \times N$ variance–covariance (vcv) matrix $\mathbf{C}$, where N is equal to the number of tips in $\mathcal{T}$. The elements $C_{i,j}$ are the shared path–length from the root to the most recent common ancestor of i and j (?). The diagonal elements ($i = j$) are simply the total distance from the root to the tips. For any model, we can describe a second matrix $\mathbf{\Sigma}$, which is the expected vcv matrix between observations at the tips. For example, under BM, in which variation accumulates proportionally to time under a single rate σ^2,

$$\Sigma_{ij} = \sigma^2 C_{ij} \tag{2}$$

and thus

$$\mathbf{\Sigma} = \sigma^2 \mathbf{C}. \tag{3}$$

11

Under a single–optimum OU model of trait evolution,

$$\Sigma_{ij} = \frac{\sigma^2}{\alpha} \exp[-2\alpha(T - C_{ij})](1 - \exp[-2\alpha C_{ij}]) \tag{4}$$

where σ^2 is the BM rate parameter, α is the strength of attraction towards to mean and T is the total depth of the tree (Hansen, 1997; Butler and King, 2004). A vcv matrix $\mathbf{U}$, which describes the unit tree is then by definition equivalent to $\mathbf{\Sigma}$ for any model. We note that in practice we are using the estimated paramters $\hat{\Theta}$ from fitting the model to construct $\mathbf{U}$. For the case of using a model with a single mean, we can transform $\mathbf{T}$ to $\mathbf{U}$ by... If we use a model with multiple means, such as a multi–optimum OU model (Butler and King, 2004; Beaulieu et al., 2012), we must

Parametric bootstrapping

Posterior predictive simulations

Phylogenetic regression

Simulations

Empirical analyses

Phylogenetic tree and Trait Data

We used a "megaphylogeny" of Angiosperms from a recent study by **?**. The tree contains 32,xxx taxa and covers xx % of familial diversity and yy % of generic diversity across all Angiosperms. We will not provide the full details

on the phylogeny here and refer interested readers to the original publication
(**?**). But in brief, the data matrix was constructed using previoulsy published
data on GENBANK, which was obtained using the program PHLAWD (**?**). 5 genes
were

For the purposes of this study, we used the MLE point estimate of the phylogeny. We could have used a set of bootstrapped trees and ran our analyses
across all of them, but for the purposes of the present study, it should not
qualitatively affect our results. This tree is published on TREEBASE (assession
number), DRYAD (accession number), and is included as part of the OPEN TREE
OF LIFE project (**?**).

Adequacy of models for plant functional traits

Results

Simulation results

Adequacy of models for plant functional traits

Discussion

13

A note on implementation

MOVE THIS ALL TO A TABLE

This approach has been implemented in the `R` package `GEIGER` (version 2.XX **?**). Currently, researchers can fit models with ML using the functions `fitContinuous` in `GEIGER` (**?**) and `OUwie` (Beaulieu et al., 2012). Researchers can also fit trait models using a Bayesian MCMC with the `GEIGER` functions `fitContinuousMCMC` (includes fossils; **??**), `rjmcmc.bm` (Eastman et al., 2011; **?**), `BayOU` (**?**) and the generalized MCMC function `phyTraitMCMC` (desribed above and in the Supplementary Online Material of this paper.) For phylogenetic regression analyses, our approach can be used in conjunction with the `gls` function in `nlme` (**?**) and using the correlation structure of `ape` (**?**). All source code for this project at `www.github.com/mwpennell/modeladequacy`.

Concluding remarks

Acknowledgments

References

Akaike, H. 1974. A new look at the statistical model identification. Automatic Control, IEEE Transactions on 19:716 – 723.

Beaulieu, J. M., D.-C. Jhwueng, C. Boettiger, and B. C. O'Meara. 2012. Modeling stabilizing selection: Expanding the ornstein-uhlenbeck model of adaptive evolution. Evolution 66:2369–2383.

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

Boettiger, C., G. Coop, and P. Ralph. 2012. Is your phylogeny informative? measuring the power of comparative methods. Evolution Page in press.

Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. The American Naturalist 164:683–695.

Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65:3578–3589.

Edwards, A. W. F. and L. L. Cavalli-Sforza. 1964. Phenetic and phylogenetic classification chap. pages 67-76. Systematics Association Publication No 6 Systematics Association, London.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees

from continuous characters. American Journal of Human Genetics 25:471–492.

Felsenstein, J. 1985. Phylogenies and the comparative method. The American Naturalist 125:1–15.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:pp. 1341–1351.

Hansen, T. F. and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of inter-specific data. Evolution 50:1404–1417.

Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. Bryan Jennings, K. H. Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, A. Purvis, R. E. Ricklefs, D. Schluter, J. A. Schulte II, O. Seehausen, B. L. Sidlauskas, O. Torres-Carvajal, J. T. Weir, and A. Ø. Mooers. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396.

Hunt, G. 2006. Fitting and comparing models of phyletic evolution: random walks and beyond. Paleobiology 32:578–601.

Hunt, G. 2012. Measuring rates of phenotypic evolution and the inseparability of tempo and mode. Paleobiology 38:351–373.

O'Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

16

Pagel, M. 1997. Inferring evolutionary processes from phylogenies. Zoologica
Scripta 26:331–348.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statisitcs
6:461 − 464.

Slater, G. J. in revision. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the cretaceous-paleogene boundary. Methods
in Ecology and Evolution .

Slater, G. J. and M. W. Pennell. in revision. Robust regression and posterior
predictive simulation increase power to detect early bursts of trait evolution.
Systematic Biology .

Thompson, E. A. 1975. Human Evolutionary Trees. Cambridge University
Press.