

# 1 Inferring Adaptive Landscapes from Phylogenetic Trees: 2 A Dissertation Proposal

3 Carl Boettiger<sup>\*,a</sup>

4 <sup>a</sup>*Center for Population Biology, University of California, Davis, United States*

---

## 5 Abstract

Phylogenetically based comparative methods are an established and rapidly expanding area of research in macroevolution. Existing approaches may produce misleading results when the traits under consideration reflect different niche specialization across the taxa in question. I propose a method that addresses these difficulties and extends our ability to ask new questions using phylogenetic comparative data, such as the inferring number of niches represented in the data and rate of evolutionary transitions between niches.

6 *Key words:* Evolution, Phylogenetics, Comparative Methods

---

## 7 1. Multiple niches can confound existing methods

8 Imagine we have identified a continuously valued phenotypic trait which seems to play an important  
9 role in determining the niches of species across some array of taxa. We seek to explain the diversity  
10 across these taxa through an adaptive radiation facilitated by adaptation in this functional trait. For  
11 instance, island Anolis lizards are thought to differ in hind limb sizes and consequently select different  
12 size perches – dividing them into different classes or ecomorphs that has been observed repeatedly in  
13 different islands (Williams, 1969). To characterize the radiation, we may wish to reconstruct ancestral  
14 states of the phenotype of this trait, estimate the rate of diversification in this trait, and explore how the  
15 diversification rate may differ between clades or have changed over time. To address these questions,  
16 we must consider both the constraints of evolutionary history as well as adaptation (Losos, 1996).  
17 Using DNA sequence data, we can construct an ultrametric phylogenetic tree showing the evolutionary  
18 connections between each of his lizard species, as in Fig. 1. Then we approach these questions using  
19 the tools of phylogenetic comparative methods.

20 Ancestral state reconstruction methods for continuous traits (Martins and Hansen, 1997; Schluter  
21 et al., 1997) are all essentially based on the Brownian motion (BM) model of character trait evolu-  
22 tion Felsenstein (1985), though in principle could be carried out with the Ornstead-Uhlenbeck (OU)  
23 model described by Hansen (1997). Using these methods on this data, one will always infer interme-  
24 diate states such as seen in Fig. 1(a). This raises some cause for concern as the predicted values have  
25 little overlap with the values observed in the present day, Fig. 1(b). This difficulty arises whenever the  
26 observed taxa represent two distinct ranges in the continuous trait, as may be expected if the trait is  
27 responsible for niche differentiation. While such limitations have been identified, (*e.g.* Losos (1999)  
28 or Cunningham (1998)), it remains challenging to account for niche differences with existing methods.

29 Estimates of diversification rate are similarly challenged by this data. The actual estimations are  
30 straight-forward: For a unit-length tree shown in Fig. 1(a) under BM, the maximum likelihood estimate  
31 for the diversification rate is  $\sigma = 9.4$ , while under an OU model the diversification rate is estimated  
32 at  $\sigma = 41$  while the stabilizing selection strength  $\alpha = 42$ , centered on a trait value of  $\theta = 7.3$ . The  
33 OU model does slightly better in model comparison measures, *i.e.* an AIC of 58 for OU vs 64 under

---

\*Corresponding author.

Email address: cboettig@ucdavis.edu (Carl Boettiger)

34 BM. This analysis prefers a model with an adaptive peak located in the middle of the valley between  
 35 the two modes of the data, resembling the same pattern as in Fig. 1(b). The stabilizing selection is  
 36 centered at the intermediate value  $\theta = 7.3$  with an equilibrium width of  $\sqrt{\frac{\sigma^2}{2\alpha}} = 4.4$ , approximately  
 37 the same as the width of either mode. Once again, the apparent multiple niche structure of the data  
 38 seems at odds with existing approaches.

39 A proposed signature of adaptive radiations is a decelerating rate of character evolution (Gavrilets  
 40 and Vose, 2005; Gavrilets and Losos, 2009), which comparative methods approaches may attempt to  
 41 measure through standardized contrasts (Freckleton and Harvey, 2006). Standardized contrasts are  
 42 constructed by differences between ancestral and descendant states weighted by the time separating  
 43 them. Applying this to the data in Fig. 1(a), we appear to see a clear signal of *accelerating* rates of  
 44 evolution, where most trait change occurs in a burst of evolution at the tips. Clearly this could just be  
 45 an artifact of the model due to the ancestral state estimates.

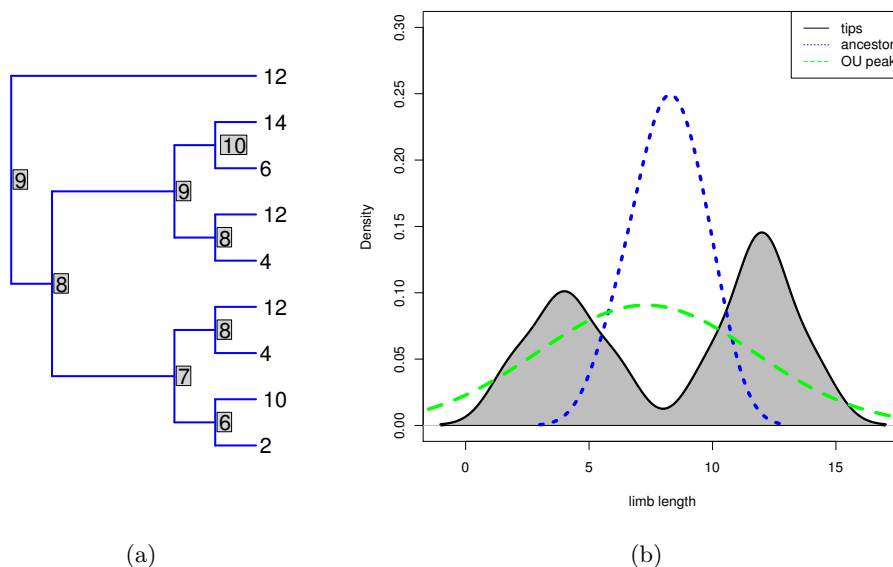


Figure 1: (a) Hypothetical Anolis hind limb lengths on a phylogenetic tree. Ancestral states inferred under Felsenstein’s Brownian motion model of evolution are shown at the interior nodes. (b) Distribution of observed limb lengths at tips compared to the inferred ancestral states

46 The challenges introduced by phylogenies that span traits sampled from different niches requires a  
 47 new approach to accurately account for the importance of phylogenetic relationships in using compar-  
 48 ative data. In this, we follow an emerging pattern in the field of comparative methods. (1) Felsenstein  
 49 (1985) first introduced BM as a model of trait evolution to enable existing applications of compar-  
 50 ative methods, usually correlations between two traits or trait and environment, to account for the  
 51 biological reality of phylogenetic relatedness. This approach was quickly adapted to answer novel ques-  
 52 tions such as estimating rates of diversification (Garland, 1992) and comparing these rates between  
 53 clades (O’Meara et al., 2006; Collar et al., 2005). While the BM model provided a reasonable de-  
 54 scription of diversification, it could not capture the biological reality of stabilizing selection, and could  
 55 consequently give spurious results when applied to traits that were strongly conserved over a phylogeny.  
 56 (2) To address this, Hansen (1997) introduced the OU model as a description of adaptive trait evolution  
 57 under stabilizing selection, which in turn introduced a new class of questions – identifying the strength

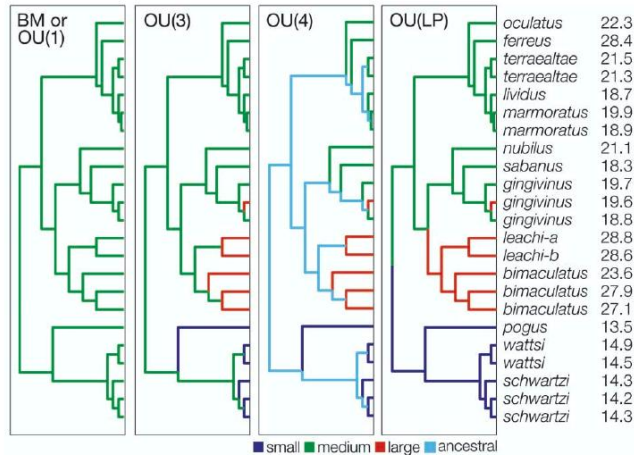


Figure 2: Possible paintings of a phylogenetic tree, taken from [Butler and King \(2004\)](#).

58 of selection and comparing model fits between OU and BM models. The existence of multiple niches  
 59 continues this evolutionary cycle. To accurately apply these existing methods on such data, we need  
 60 an approach that can capture multiple niches, impossible with *linear* models such as OU and BM.  
 61 (3) Creating such a method not only allows us to accommodate this biological reality, but opens the  
 62 door to once again exploring new questions in a comparative context, such as how many niches may  
 63 be represented and how frequently transitions have occurred between them.

64 There is an existing approach attempting to deal with the challenge of multiple niches known  
 65 as *painting*, which applies different OU models to different branches on the tree ([Butler and King,](#)  
 66 [2004](#)). In this model-comparison approach one asks if there is evidence that a particular clade has a  
 67 statistically significantly different optima than other clades (or branches) on the tree; much as has been  
 68 done for BM model in [O’Meara et al. \(2006\)](#). There are several difficulties with this approach. First,  
 69 it requires an informed way of selecting possible paintings of the tree, identifying different branches  
 70 with different colors indicating they are governed by an independent evolutionary regime, as in Fig 2.  
 71 Searching over all possible paintings may not only be infeasible but uninformative, as the transitions  
 72 between regimes remove the phylogenetic signal. While the interpretation of different parts of the  
 73 tree falling under different selective regimes seems very reasonable, transitions between such regimes  
 74 may not occur only at nodes, particularly those nodes where both branches have a surviving ancestor  
 75 (extinct lines introduce additional nodes on the tree). Further, it is difficult to estimate transition  
 76 rates between regimes or the number of regimes that best fits the data – both potential quantities of  
 77 interest. By taking a much more general, *nonlinear* model framework we will be able to address each  
 78 of these challenges.

79 The introduction of OU model and the painting of multiple OU models onto a tree has allowed the  
 80 field to enter the realm of model testing and model comparison. These comparisons are typically based  
 81 on information criteria such as AIC, which penalize models that have more parameters by a certain  
 82 factor. However, there is little reason to believe that AIC is fundamentally meaningful way to perform  
 83 this discrimination. Using the data of [Butler and King \(2004\)](#) I fitted a Brownian motion model which  
 84 I used to generate 1000 simulated datasets, then repeated the analysis on the simulated data. In 391  
 85 of the sets the best chosen model of [Butler and King \(2004\)](#), which uses three different OU models,  
 86 still has the best AIC score, indicating the penalization for extra parameters is not sufficient. As can  
 87 be seen in figure 3, this 4:6 error rate is little better than a coin flip to choose models. Clearly the

88 question of model selection must be considered more carefully.

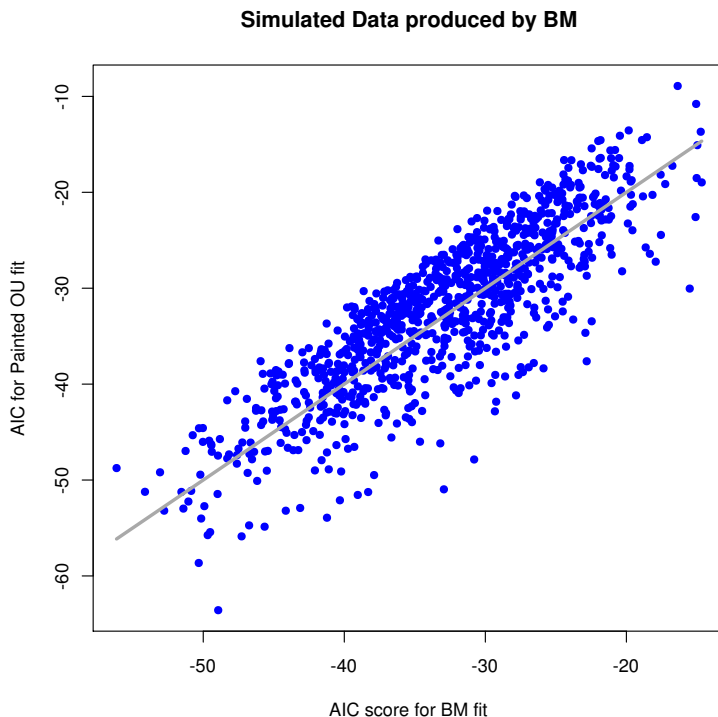


Figure 3: Points falling below the 1:1 line indicate a better (smaller, more negative) AIC score for the painted OU model despite the fact that the data was generated under BM. Note that some of the widest differences in AIC are errors.

89 In light of this, we need a new method. I propose to develop a robust, general Bayesian approach  
90 that avoids these difficulties. Such a framework could explicitly account for multiple niches, and identify  
91 clusters in traits or behaviors or ecology that representing niche differentiation. In the process we will  
92 also visit how this approach can be used to capture other nonlinear models as well, such as bounded  
93 Brownian motion models or even arbitrarily general models.

## 94 2. A model for multiple niches

95 Adaptive radiations are characterized by extensive diversification into a variety of ecological niches  
96 over relatively short time scales that may generate much of the diversity of life. Typically this may be  
97 driven either by invading an underutilized environment or through the development of a key innovation  
98 (such as flight) which makes an array of possible niches available (though both these concepts of an  
99 empty niche and a key innovation are subjects of active research).

100 Few examples are better studied than that of the Anolis lizards, particularly the six recognized  
101 ecomorphs of the Greater Antilles Islands in the Caribbean. This radiation appears to have repeated  
102 itself on four islands, through a balance of repeated invasions and repeated evolution. The ecomorphs  
103 represent clusters of a suite of ecological, morphological and behavioral traits that occur repeatedly  
104 though through different species across the islands, (see [Losos \(2009\)](#) for a thorough discussion).  
105 Underlying the description of ecomorph is an *evolutionary* assumption – that evolution acts on this  
106 suite of traits to generate the repeated emergence of the clusters.

- 107 1. It matters which traits we include in the description. We are interested in describing functional  
108 traits. Measuring arbitrary traits and clustering along PCA is insufficient, as the functional trait  
109 may not be a linear combination of the morphology – indeed the “many-to-one” mapping that  
110 results from nonlinear interactions between morphology and performance is thought to be both  
111 pervasive and important in our understanding of evolution. Fish jaw morphology provides an  
112 excellent example, where performance can be well-characterized by suction index of a fish – an  
113 example of the “right” functional trait which is not a linear combination of basic morphological  
114 values.
- 115 2. We would like a robust quantitative way to define clusters along these trait axes. We should be  
116 able to determine how many clusters exist and quantify how well we can discriminate between  
117 peaks. Having identified interesting traits to test, there are two broad reasons why this clustering  
118 may not be immediately apparent – noise and phylogenetic inertia. Noise could come from  
119 many sources, from measurement error to environmental variation to the suite of other selective  
120 demands on the traits; all of which serve to blur out distinct clusters. Phylogenetic inertia may  
121 blur the distinction between groups if certain species have actively evolving in the direction of a  
122 particular cluster from the vicinity of another, and hence may not fall clearly into one group.
- 123 3. We would like to do so in a way which reflects the evolutionary process. This will allow us to  
124 better interpret the evolutionary significance of the parameters which define the clusters – for  
125 instance, strength of selection within groups and rates of transitions between groups.
- 126 4. We would like to utilize phylogenetic information to inform the clusters. If we know the phyloge-  
127 netic tree connecting all species in the analysis, we may be able to use this information not only  
128 to correct for phylogenetic inertia but to penetrate the noise that may obscure clusters.

129 To make all of this more precise, we define a model that captures each of these elements. The  
130 model consists of  $n$  regimes over trait space of dimension  $k$ . Within each regime is a single, linear,  
131 multidimensional attractor driven by a white noise perturbation – *e.g.* the multi-dimensional Ornstead-  
132 Uhlenbeck model,

$$dX = \alpha(\Theta - X)dt + \sigma dW_t \tag{1}$$

133 While in general  $X$  and  $\theta$  are  $k$  dimensional trait vectors,  $\alpha$  and  $\sigma$   $k \times k$  matrices of the trait  
134 correlations and covariances and  $dW_t$  a  $k$  dimensional Weiner process (Brownian walk). For simplicity,  
135 we will often consider a single trait,  $k = 1$ . In addition, a trait value may take a large jump into a  
136 new regime – representing a sudden transition of environment or other selective pressure. Each regime  
137 may have a unique transition rate into every other regime, reflecting how common available niches may  
138 be, which we take to be constant. Assume that these transitions are independent of each other, we  
139 have a Poisson process we can represent by the transition rate from regime  $i$  to  $j$  as the  $i, j$  term in  
140 transition matrix  $\mathbb{Q}$ ,  $i \neq j$ , where the diagonal terms are such that rows sum to zero. Fig. 4 provides  
141 a conceptual illustration of such a model. For simplicity, we may assume that all transition rates are  
142 symmetric (the rate from  $i$  to  $j$  equals the rate from  $j$  to  $i$ ) or that some transitions are impossible.  
143 Another convenient simplification may be that all regimes are characterized by the same diversification  
144 rate  $\sigma$ , making the width determined entirely by  $\alpha$  (recall that at the stationary state the distribution  
145 is Gaussian with variance  $\frac{\sigma^2}{2\alpha}$ ).

146 Such a model can capture each of the criteria discussed above.<sup>1</sup> (1) This model is defined over the  
147 trait space of interest, or a subspace thereof. (2) It provides a quantitative description the clusters:

---

<sup>1</sup> It is worth noting that this is certainly not the only formulation of a model that satisfies these objectives. For instance, we could have used a singular *nonlinear* stochastic differential equation rather than this collection of multiple linear ones (the OU models). The regime formulation has both conceptual and practical advantages, such as separating

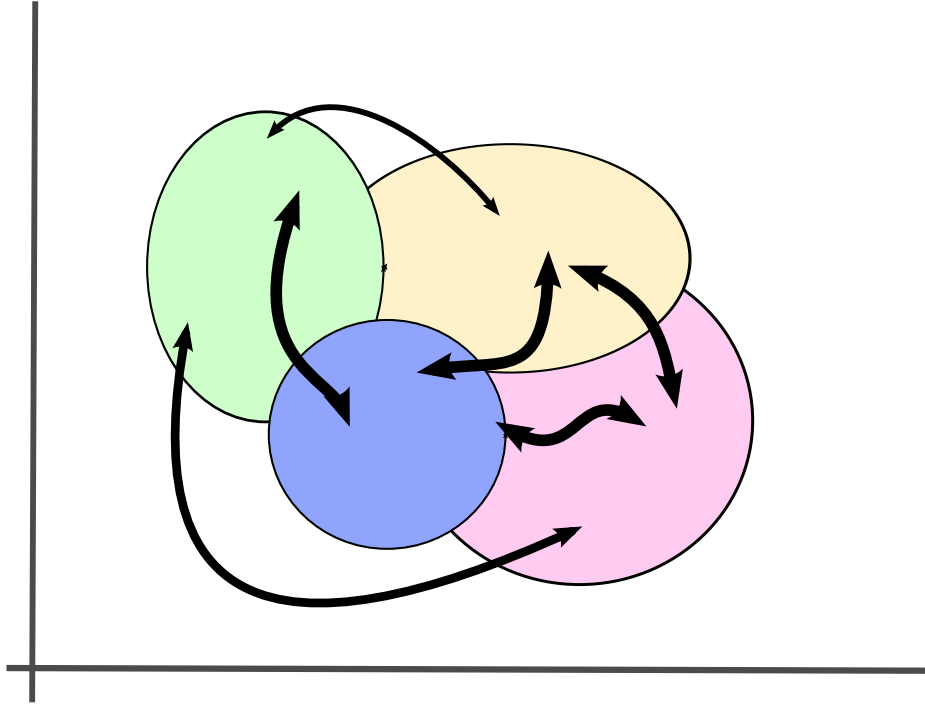


Figure 4: A conceptual model of four ecomorphs and transitions between them. Axes represent trait space, and ecomorphs differ in the location of their optima with respect to each trait and the strength of selection in each dimension. Rates of transitions may differ (as indicated by different weights) and need not be symmetric. Mathematically this could be captured by an OU process centered at the optimum of each cluster with the appropriate strength of selection  $\alpha$  relative to diversification  $\sigma$ . Transitions could be captured as a Poisson process with transition rate matrix  $\mathbb{Q}$ .

148 the number ( $n$ ), size ( $\frac{\sigma_i^2}{2\alpha_i}$ ), and locations ( $\theta_i$ ) and capacity (through the transition rates – regimes  
 149 with rapid transitions out will represent niches with lower occupancy). (3) The model has a meaningful  
 150 evolutionary interpretation –  $\theta_i$  are selective optima under  $i$  different environments/interactions,  $\alpha_i$  the  
 151 strengths of selection, variation within a regime is captured by  $\sigma_i$  while diversification rate across niches  
 152 is determined by the transition rates. This can explicitly model the noise due to competing selective  
 153 forces, or other fluctuations through  $\sigma_i$ . Lastly, it is a dynamic model, explicitly describing a time  
 154 dependent process. This enables us to (4) use the phylogenetic information in the tree by assuming  
 155 that separate branches of the tree spawn independent instances of the process starting from the same  
 156 initial condition. Not only will this allow us to correct for phylogenetic inertia, but it improves our  
 157 ability to estimate dynamic quantities of the model.

### 158 3. Method

#### 159 3.1. The right data

160 The data refers to a suite (vector) of observed traits across taxa their ultrametric tree. We will  
 161 assume for the moment that the traits are continuously valued. We imagine defining the observed

---

the process of diversification within a regime from that between regimes, while the linearity makes the regimes easy to treat.

162 traits at all of the tips of the tree, where they are interpreted as mean trait values for the species  
163 (though in practice they may consist of single samples). This data could be morphological (limb length),  
164 behavioral (movement speed) or ecological (structural habitat). For the moment we will ignore both any  
165 uncertainty in the tree and variation in around these trait means, though later we will explore including  
166 both. This kind of data and assumptions are typical of comparative methods applications. While it  
167 is common to transform trait values and to use principle component analysis (PCA) combinations of  
168 trait values, we caution against the unconsidered application of either of these, and prefer to focus  
169 on a carefully selected, interpretable trait. Log-transforms are particularly common as a way of non-  
170 dimensionalizing the data. The difficulty in doing so is that the expectation of the transformed trait  
171 is not the same as the transformation of the expected trait value under non-linear transformation  
172 such as logarithms – *e.g.* if the expected log limb size is constant, then the expected limb size is  
173 shrinking. As mentioned before, the difficulty in PCA also stems from the linearity assumption. A good  
174 example is the concept of suction index in fish feeding performance. Suction index is a non-dimensional  
175 morphological trait that is closely correlated with feeding performance and differs significantly across  
176 fish with different diets. The index is a nonlinear function of five morphological measurements of  
177 the fish jaw, and is likely to be much more informative than generic transforms. Starting with good  
178 functional morphology greatly enhances the potential of these approaches.

### 179 3.2. The Joint Probability

180 Having framed the model, we must define the how to evaluate the joint probability of the observed  
181 data given the model. This joint probability will be the fundamental building block of the method  
182 which we will use to fit parameters of the model and select between models. The joint probability of  
183 seeing the observed trait data at the particular nodes at which it appears under a given model can be  
184 calculated by unfolding the tree and integrating over the unknown nodes as follows: If we know the  
185 root, we can use the transition densities to determine probability distribution of trait values for each  
186 of its daughter nodes by knowing the length of time between them,  $w(x_r, x_i, t_{ij})$ . As we don't know  
187 the root, we must integrate this transition density over all possible values of the root node,  $P(x_r)$ .  
188 Similarly, we must integrate over possible values for each of the internal nodes, as they are unknown.  
189 To visualize this, consider a simple example such as the tree given in Fig. 5. We could write the joint  
190 probability as:

$$\begin{aligned}
 P(x_1, x_2, x_3, x_4) = \int dx_7 \left[ \left[ \int dx_5 w(x_7, x_5, t_5) w(x_5, x_1, t_1) w(x_5, x_2, t_1) \right] \right. \\
 \left. \times \left[ \int dx_6 w(x_7, x_6, t_6) w(x_6, x_3, t_3) w(x_6, x_4, t_3) \right] P(x_7) \right] \quad (2)
 \end{aligned}$$

191 In a numerical implementation, this can be defined as a recursion over matrix multiplications.  
192 Assume the trait assumes values in an interval  $\mathbb{I}$  which we discretize into  $m$  points. Then along each  
193 branch we can place the transition rate as an  $m$  by  $m$  matrix of the transition rates from any point in  
194 the space to any other during that time interval. As the values at the tips are known, we need only the  
195 column corresponding to transitions to the observed state. If the root node is specified (as a parameter  
196 of the model) then its branches are row vectors from the given trait to each point along the interval.  
197 The recursion starts at the root with this row vector and multiplies the row vector it receives from  
198 the previous node by the transition matrices of its left and right children. This proceeds up the tree  
199 until it reaches a tip, in which case the row vector multiplies the column vector of the tip to produce  
200 a scalar. The product of all these scalars is the joint probability.



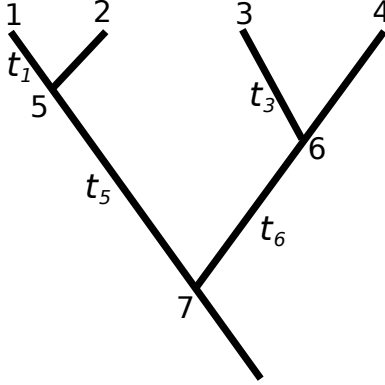


Figure 5: A simple phylogenetic tree.

201 If we assume that when a population transitions to a new state it immediately assumes trait values  
 202 from the stationary distribution of the process, then the transition probability along a branch from  
 203 node  $i$  to  $j$  is given by:

$$w(x_i, s_i, x_j, s_j, t_{ij}) = \underbrace{\left[ e^{t_{ij} Q} \right]_{s_i s_j}}_{\text{all transitions}} \Pi_{s_j}(x_j) + \underbrace{\mathbb{I}_{s_i=s_j}}_{\text{Indicator fn}} \left[ \underbrace{-e^{t_{ij} Q_{s_i s_i}} \Pi_{s_i}(x_j)}_{\text{minus case no transitions}} + \underbrace{e^{t_{ij} Q_{s_i s_i}} w_{s_i}(x_i \rightarrow x_j, t_{ij})}_{\text{use the transition density instead}} \right] \quad (3)$$

204 where  $\Pi_{s_j}(x_j)$  is the stationary distribution of regime  $s_j$  and  $w_{s_i}(x_i \rightarrow x_j, t_{ij})$  is the time dependent  
 205 rate of the OU process (see Appendix).

#### 206 4. Fitting the model

207 Once we can calculate the joint probability of the data under a given model and parameterization,  
 208 we have a wide array of potential methods that will help us search for the best value of the parameters  
 209 and subsequent selection between models. For a variety of reasons, it is particularly appealing to frame  
 210 this question in Bayesian context where the analysis will be performed through Markov Chain Monte  
 211 Carlo (MCMC). Of particular interest is the trans-dimensional MCMC proposed by [Green \(1995, 2003\)](#),  
 212 which provides a unified framework for simultaneously fitting model parameters and selecting between  
 213 models that differ in the number of regimes (and hence the number of parameters) they have. Without  
 214 such approach the comparison of number of transitions can be done in an information-theoretic context  
 215 such as AIC or Bayes (Schwartz) Information Criterion after the models have been fitted.

#### 216 5. Example Systems

217 Labrid fishes are a potentially rich example of such a radiation, accounting for over 600 species of  
 218 wrasses and parrotfish found in diverse micro-habitats in coral reefs around the world and spanning a  
 219 particularly impressive range of dietary niches, including molluscs, corals, fish parasites, zooplankton  
 220 and other fishes ([Bellwood et al., 2006](#)). Dietary preference is strongly controlled by both skull and  
 221 pectoral fin morphology which limits the kind of resources the fish can capture and process, and the  
 222 morphological diversity in these traits reflects the ecological variation in diet ([Collar et al., 2008](#)).



223 Extensive research has documented the particular morphological traits responsible for feeding per-  
224 formance. Labrid fish swim almost entirely with their pectoral fins alone, rather than relying on body  
225 and caudal fin motion. Fins vary between high aspect ratio (AR), long, thin shapes that flap in a  
226 wing like motion, and low AR, rounder shapes that row in a paddle motion, Fig. 6(a). High AR fins  
227 perform at higher speeds and these species tend to occupy higher flow areas of the reef, while lower  
228 AR fins perform better at rapid turns and maneuverability (Bellwood and Wainwright, 2001). Jaw  
229 morphology differs between strong sucking jaws and powerful crushing jaws, Fig 6(b) which correlate  
230 strongly with diet composition (Wainwright, 1988). As the functional roles of these complex traits  
231 has clear implications for the array of possible niche types characteristic in the diversity of Labrids,  
232 they provide a natural system to test how phylogenetic comparisons inform our understanding of niche  
233 differences. Using these traits and available ultrametric trees for Labrid fishes, we want to explore the  
234 following questions:

- 235 1. How many different niches are represented?
- 236 2. How frequently do transitions occur between niche strategies?
- 237 3. Where is each niche optima located, how far apart and with what width?
- 238 4. Can we make niche assignments of present-day species based on this reconstruction?

239 These traits may capture much of the functional diversity responsible for the niche diversity that  
240 may underly an adaptive radiation, and phylogenetic comparative methods have already proved useful  
241 in exploring their evolutionary significance (Collar et al., 2005, 2009, 2008). As they may underlie niche  
242 differentiation, they may exhibit the kinds of clustering shown to be a challenge for existing methods  
243 when exploring concepts such as ancestral states or changing rates of diversification. However, due  
244 to competing demands of selection on these and other correlated traits as well as other sources of  
245 variability, the distribution of traits observed in the present day may not reveal the obvious niche  
246 differences presented in the imaginary example of Fig. 1(b). In fact, many potentially differentiating  
247 traits often present an essentially unimodal distribution of values across observed species. Our goal,  
248 then, is use the way in which those values are distributed across a phylogenetic tree to infer how many  
249 different niche strategies may be represented across this distribution of traits. To see how this might  
250 work in principle, let us consider another hypothetical example.

251 The first step in demonstrating that such an inference is possible in principle is to demonstrate  
252 how a continuous trait under stabilizing selection for one of two distinct optima representing different  
253 niche strategies could produce a unimodal distribution in trait values due to random selection on other  
254 traits, environmental variation, or genetic drift. We illustrate this with a simulation shown in Fig. 7  
255 under different intensities of such random effects.

256 Having seen how the problem arises, we describe how the phylogeny can help us reveal the underlying  
257 structure despite the noise. Given some unimodal distribution of traits, there are many possible  
258 arbitrary phylogenetic trees we could draw potentially showing how they are related. The simplest  
259 description of the unimodal peak will be a single attractor (OU) at the mode or BM diffusion that  
260 began at that mode. The key to distinguishing between BM, OU, and multiple niche description will  
261 rely on the differences in closely related species compared to those in distantly related species. Under  
262 BM, the distance in traits should be increase (as the square root) of the branch lengths between traits  
263 (as this model has no stationary distribution, distances always increase). Under OU model, distantly  
264 related species should appear independent, given by the decorrelation time introduced by the selective  
265 force ( $\alpha$ ), since they become samples from the stationary distribution of the model. In a multiple niche  
266 model, distantly related species will still appear more correlated than expected, as they may still be  
267 trapped near the same niche, while not having reached the stationary distribution where they may be  
268 independently likely to be found in any of the niches. We simply require an approach that can consider

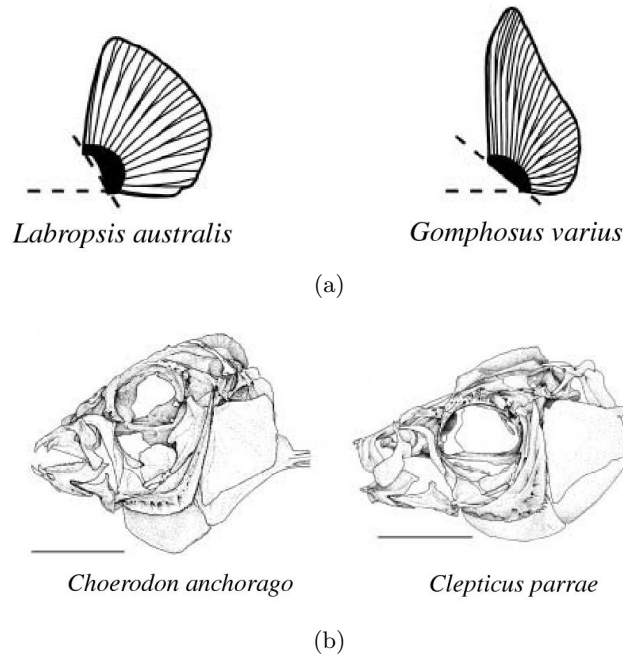


Figure 6: Fin morphology falls into essentially two distinct classes: fins with low aspect ratio (left) for faster turns and fins with high aspect ratio (right) for sustained swimming at high speed, reproduced from (Collar et al., 2008).

269 the impact of the phylogenetic structure has on the likelihood we observe the particular distribution  
 270 of data along the tree tips.

271 Ultimately, we would like such a method to be implemented as an open source and freely distributed  
 272 software package to facilitate this analysis, as researchers have done for existing methods *i.e.* Harmon  
 273 et al. (2008); Butler and King (2004); Hansen et al. (2008); Pagel and Meade (2006); Paradis (2004);  
 274 O’Meara et al. (2006).

## 275 6. Extensions

276 Beyond the ability to search for number and positions of niches and transitions between them,  
 277 this general framework allows many other possible extensions to the phylogenetic comparative method  
 278 framework. To convey this potential, we briefly list several possible directions here which we hope to  
 279 incorporate.

### 280 6.1. Statistical indicators: estimating power, confidence intervals and posteriors

281 Some statistical validation should be part of any method. Particularly important here is the ob-  
 282 servation that any particular outcome is intrinsically unlikely, given the enormous number of possible  
 283 outcomes of any model. Nevertheless, estimates of power are relatively straight forward, if computa-  
 284 tionally intense, to provide. Given a selected model with its maximum likelihood parameter values, one  
 285 can create a set of simulated data sets and then run the method on this simulated data and compare  
 286 the spread in parameter estimates from each of these runs. Creating simulated data sets is quite fast,  
 287 though clearly the comparison will scale linearly with the number of simulated sets used. In the same  
 288 fashion one can estimate the power of the model selection algorithm by asking for how many of the  
 289 simulated data sets does the method select the same model as the best choice?

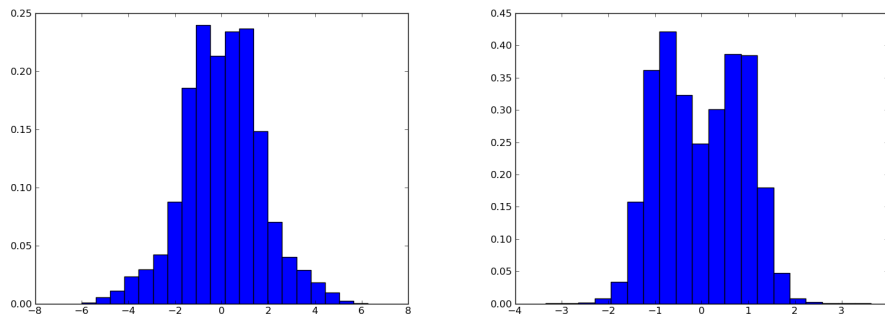


Figure 7: Variance due to random effects obscures the underlying two basins of attraction (niches) on the left. On the right, under reduced variance the shape of the adaptive landscape is more visible

## 290 6.2. Multi-dimensional traits

291 Though we have discussed this in the context of a single continuous trait, the principles involved  
 292 are the same for vector valued traits. This requires determining the transition density from any point  
 293 in this vector space to any other point, which quickly becomes computationally intensive in the general  
 294 case. In practice, several analytical shortcuts may be possible. Further, if the evolution of the traits is  
 295 independent, a scalar valued model could be applied independently for each trait. Such an approach  
 296 could be used to create independent contrasts under models other than BM that could then be tested  
 297 for correlations with classical regression techniques.

## 298 6.3. Discrete traits

299 While we have focused on continuous traits for simplicity of presentation, this method is equally well  
 300 suited to handle discrete phenotypic traits. While Pagel (1994); Pagel et al. (2004) already provide  
 301 a phylogenetic comparative methods framework for discrete traits, our approach can handle both  
 302 simultaneously; in fact the regime model is one such example of this – the regimes could represent  
 303 discrete traits which influence which optimum a continuous trait centers around; *i.e.* nocturnal /  
 304 diurnal behavior and temperature preference, or carnivory / omnivory / herbivory and a continuous  
 305 measure of tooth morphology.

## 306 6.4. Incorporating fossils

307 Incorporating fossil information in this approach is very straight forward, as the information can  
 308 be specified at the appropriate node and then treated like a tip value, rather than being integrated  
 309 over. Available fossil information can also be used to assess the accuracy of ancestral state reconstruc-  
 310 tion under this approach by running the model without the fossil data and comparing. The method  
 311 could also be applied to evolutionary experiments in microbes, where freezers could preserve samples  
 312 throughout the evolutionary history for comparison.

## 313 6.5. Changing Environments: Niche creation and loss

314 This relaxes the Markov assumption in the transition densities, calculating transition densities for  
 315 going from state  $x_1$  at time  $t_1$  to state  $x_2$  at time  $t_2$  ( $w(x_1, x_2, t_1, t_2)$ ), rather than going from  $x_1$  to  $x_2$   
 316 in an interval  $t = t_2 - t_1$ . Simple applications of this would simply divide the tree into epochs – say after  
 317 a fixed time (perhaps predetermined based on external data) transition densities are computed with a  
 318 model of three peaks and before that time with a model using only two peaks, and this fit is compared  
 319 to the model scores of always using two or always using three peaks. This provides a potential test

320 of whether the data shows any evidence of such a shift occurring, or at least an estimate that the  
321 data lacks the power to detect such a change. This can be used as a way of incorporating additional  
322 information about climate such as might be estimated from timing of ice ages or the introduction of  
323 a novel species or an antibiotic (say, for model of bacterial evolution) into the system, or attempt to  
324 estimate the timing of the transition as a model parameter.

325 The goal of the extensions is to allow more of what we know about the biology and its environmental  
326 context to inform our method, rather than increase the number of things we attempt to estimate with  
327 limited data. This is crucial to bear in mind, for while they may sound different, the ability to do one  
328 enables the ability to do the other – adding a source of information or estimating it as a free parameter.  
329 Thus adding fossils is the converse of estimate ancestral states, and estimating such a time shift the  
330 converse of using external information about such a change to enforce a shift. While it is tempting  
331 to be as agnostic as possible, the goal should be more on the former use, fixing constraints based on  
332 additional data, than on the latter – estimating more quantities with less data.

### 333 6.6. Bounded diversification and noise models

334 Brownian motion may provide a reasonable model within a certain range of trait values but be  
335 bounded by certain constraints. We may imagine this as a random walk in a box, and attempt to  
336 estimate where the locations of the walls of this box might be based on the data. These could be  
337 represented by reflecting boundaries or as soft boundaries where species experience stronger selection  
338 towards the center if they approach the boundary but no directional selection when far from any wall.

339 An alternative construction to this might simply limit the phenotypic variance available at the  
340 extreme ends of a range. For instance, one could choose a model of the form:

$$dX = \beta \exp\left(\frac{-(X - \theta)^2}{2\sigma^2}\right) dB_t, \quad (4)$$

341 a random walk where the size of the steps decreases away from the center of the range  $\theta$ , that is, high  
342 diversification rates are found near the center of the trait distribution but decrease as one moves away.<sup>2</sup>

### 343 6.7. Ecologically realistic models

344 Ecological interactions are largely abstracted away in this approach. [Gavrilets and Vose \(2005\)](#) and  
345 [Gavrilets and Losos \(2009\)](#) consider several detailed individual-based computer simulations of adaptive  
346 radiations driven by explicit mechanisms. It would thus be a valuable test what level of detail can be  
347 recovered by the methods proposed here using data generated under these simulations (which recreate  
348 both the phylogenetic history and the character trait evolution).

349 Such individual, mechanistically based descriptions could also be employed directly in this approach.  
350 Recall that the only conditions on possible models is the ability to somehow specify (potentially by  
351 simulation) a transition density. In this way such rich mechanistic simulations can be directly tested  
352 against phylogenetic data, or used to approximate a transition density equation that can be so tested.

### 353 6.8. Scoring empirical performance landscapes

354 This approach is not limited to adaptive landscapes or niche structures that can be specified by  
355 simple functions. In some cases, researchers can actually measure performance of an organism with  
356 respect to a fitness correlate across a diversity of morphologies – such as suction index or prey capture.  
357 These performance landscapes could be input directly into the model as representations of fitness

---

<sup>2</sup>One could view this as an SDE for the transformed variable with constant noise term by Lamperti transform, see [Iacus \(2008\)](#).

358 constraints with respect to these traits, rather than the simple multi-modal functional forms otherwise  
 359 assumed. The method would not have to fit any parameters for the shape of this landscape, though it  
 360 may choose the best fit of a noise parameter, and could return likelihood score for the model. While not  
 361 directly amiable to simple model comparison metrics such as AIC, the ability to determine likelihood  
 362 alone may be informative.

## 363 7. Significance and Broader Impacts

364 An understanding of the mechanisms and patterns responsible for generating biodiversity lies at  
 365 the heart of many important challenges we face today – from vanishing biodiversity due to exploitation  
 366 and anthropogenic environmental changes and the ability of existing species to adapt in face of those  
 367 changes to the control of diseases and pests that continually evolve away from control measures. Thanks  
 368 to rapid developments in sequencing technology, bioinformatic methods and computational power,  
 369 phylogenetic information is increasingly available for a diversity of taxa. Despite there limitations in  
 370 power, simplifying assumptions and other shortcomings, phylogenetic comparative methods often offer  
 371 us our only periscope back into the past of evolutionary history, beyond the temporal capacities of  
 372 experiment and the paucity of the fossil record. New methods both extend our ability to utilize this  
 373 information and lessen our chances of being trapped by the limited tools we have available.

374 The continued development in phylogenetic comparative methods influences not only how we ap-  
 375 proach problems but in how we approach science as a process. We select the best description available  
 376 based on the observed evidence rather than evaluating and rejecting descriptions serially. We use meth-  
 377 ods customized for the particular biological questions we seek to answer, rather than one-size-fits-all  
 378 methods designed to ignore and abstract away that biology. Though the methods become more com-  
 379 plex, they are implemented in software that makes it easier for researchers to easily and accurately  
 380 implement the method in way that is repeatable, standardized, and easily verifiable by the rest of the  
 381 community. Software is open source, allowing the research community to see clearly how details and  
 382 methods are implemented, as well as catch errors or suggest improvements or extensions that can easily  
 383 be distributed. Research communities emerge around the infrastructure these methods create, such  
 384 as email lists, forums, websites and wikis that connect researchers across boundaries of institutions,  
 385 nationalities, disciplines and academic degrees. New data sets and methods are more rapidly dissemi-  
 386 nated and shared through the common infrastructure, and science proceeds at both a greater pace and  
 387 with greater cross-validation.

### 388 A. Model Library

389 Brownian Motion

$$dX_t = \sigma dB_t \quad (5)$$

390 Ornstead-Uhlenbeck

$$dX_t = \alpha(\theta - X_t)dt + \sigma dB_t \quad (6)$$

391 N Gaussian Niches:

$$dX_t = \left( \sum_i \alpha_i (\theta_i - X_t) e^{-\frac{(\theta_i - X_t)^2}{2\omega_i^2}} \right) dt + \sigma dB_t \quad (7)$$

## 392 B. Why the linear process is easy

393 If the continuous trait dynamics are generated by a diffusion process with linear rates, *i.e.* an  
394 stochastic differential equation (SDE) of the form  $dX_t = f(X_t)dt + g(X_t)dB_t$  and  $f(x)$  and  $g(x)$  are  
395 linear functions of  $x$ , then the resulting diffusion is a Gaussian process. Any collection of points resulting  
396 from a Gaussian process has a multivariate normal distribution. Further, if evolution unfolds on a  
397 phylogenetic tree according to a Gaussian process, the joint distribution of any set of taxon phenotypes  
398 is a multivariate normal, as proved in Hansen and Martins (1996). This is easily demonstrated by  
399 induction: starting with MVN set of traits  $X$  we can always add a node  $Y$  that is not an ancestor to any  
400 of the existing nodes, then the joint distribution of  $X$  and  $Y$  is multivariate normal. Consequently, to  
401 determine the joint distribution it suffices to calculate the expected trait values and variance-covariance  
402 structure of the nodes.

$$\text{Cov}(X_i, X_j) = \text{Cov}(E(X_i|X_z), E(X_j|X_z)) \quad (8)$$

403 Because  $f$  is linear, it is straightforward to find the moments of the SDE. The conditional expect-  
404 ation  $E(X_i|X_z)$  comes from solving the ODE with linear  $f$  and constant  $g$ ,  $dE(X_i) = E(f(X_i))dt$   
405 with the initial condition  $E(X_i(0)) = X_z$ . Similarly we can find the variance: for instance, for the OU  
406 process with initial condition  $X(0) = X_a$  we have,

$$E(X_t|X_a) = \alpha \int_0^t \theta(s)e^{-\alpha(t-s)}ds + X_a e^{-\alpha t}, \quad (9)$$

$$\text{Var}(X_t|X_a) = \frac{\sigma}{2\alpha}(1 - e^{-2\alpha t}), \quad (10)$$

$$\text{Cov}(X_i, X_j) = e^{-\alpha t_{ij}} \text{Var}(X_a), \quad (11)$$

409 see Hansen (1997).

410

## 411 References

- 412 Bellwood, D. R., Wainwright, P. C., 2001. Locomotion in labrid fishes: implications for habitat use  
413 and cross-shelf biogeography on the Great Barrier Reef. *Coral Reefs* 20, 139–150.
- 414 Bellwood, D. R., Wainwright, P. C., Fulton, C. J., Hoey, A. S., 2006. Functional versatility supports  
415 coral reef biodiversity. *Proceedings of The Royal Society B* 273 (1582), 101–7.  
416 URL <http://www.ncbi.nlm.nih.gov/pubmed/16519241>
- 417 Butler, M., King, A., December 2004. Phylogenetic Comparative Analysis: A Modeling Approach for  
418 Adaptive Evolution. *The American Naturalist* 164 (6), 683–695.  
419 URL <http://www.journals.uchicago.edu/doi/abs/10.1086/426002>
- 420 Collar, D. C., Near, T. J., Wainwright, P. C., August 2005. Comparative analysis of morphological  
421 diversity: does disparity accumulate at the same rate in two lineages of centrarchid fishes? *Evolution;*  
422 *international journal of organic evolution* 59 (8), 1783–94.  
423 URL <http://www.ncbi.nlm.nih.gov/pubmed/16329247>
- 424 Collar, D. C., O’Meara, B. C., Wainwright, P. C., Near, T. J., 2009. Piscivory Limits Diversification  
425 of Feeding Morphology in Centrarchid Fishes. *Evolution* 63 (6), 1557–1573.  
426 URL <http://www.bioone.org/doi/full/10.1111/j.1558-5646.2009.00626.x?prevSearch=>

- 427 Collar, D. C., Wainwright, P. C., Alfaro, M. E., 2008. Integrated diversification of locomotion and  
428 feeding in labrid fishes. *Biology Letters* 4 (1), 84.  
429 URL <http://intl-rsbl.royalsocietypublishing.org/content/4/1/84.full>
- 430 Cunningham, C., September 1998. Reconstructing ancestral character states: a critical reappraisal.  
431 *Trends in Ecology & Evolution* 13 (9), 361–366.  
432 URL <http://linkinghub.elsevier.com/retrieve/pii/S0169534798013822>
- 433 Felsenstein, J., 1985. Phylogenies and the comparative method. *American Naturalist* 125, 1–15.  
434 URL <http://www.journals.uchicago.edu/doi/abs/10.1086/284325>
- 435 Freckleton, R. P., Harvey, P. H., 2006. Detecting Non-Brownian Trait Evolution in Adaptive Radiations.  
436 *PLoS Biology* 4 (11), 2104–2111.
- 437 Garland, T. J., December 1992. Rate Tests for Phenotypic Evolution Using Phylogenetically Independent  
438 Contrasts. *The American Naturalist* 140 (3), 509 – 519.
- 439 Gavrilets, S., Losos, J. B., 2009. Adaptive radiation: contrasting theory with data. *Science* 323, 732–  
440 737.  
441 URL <http://science.samxxzy.ns02.info/cgi/content/abstract/323/5915/732>
- 442 Gavrilets, S., Vose, A., December 2005. Dynamic patterns of adaptive radiation. *Proceedings of the*  
443 *National Academy of Sciences of the United States of America* 102 (50), 18040–5.  
444 URL <http://www.ncbi.nlm.nih.gov/pubmed/16330783>
- 445 Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model deter-  
446 mination. *Biometrika* 82 (4), 711.  
447 URL <http://biomet.oxfordjournals.org/cgi/content/abstract/82/4/711>
- 448 Green, P., 2003. Trans-dimensional markov chain monte carlo. *Highly structured stochastic systems*  
449 27, 179198.  
450 URL <http://books.google.com/books?hl=en&lr=&id=NecQlPWnN3AC&oi=fnd&pg=PA179&dq=Trans-dimensional+Markov+chain+Monte+Carlo&ots=wOnwarZKQd&sig=08sciWOMmaUqRl6eEzDltzTgzLo>
- 453 Hansen, T. F., 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51 (5),  
454 13411351.  
455 URL <http://www.jstor.org/stable/2411186>
- 456 Hansen, T. F., Martins, E. P., 1996. Translating between microevolutionary process and macroevolu-  
457 tionary patterns: the correlation structure of interspecific data. *Evolution* 50 (4), 14041417.  
458 URL <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Translating+Between+Microevolutionary+Process+and+Macroevolutionary+Patterns#0>
- 460 Hansen, T. F., Pienaar, J., Orzack, S., 2008. A comparative method for studying adaptation to a  
461 randomly evolving environment. *Evolution* 62 (8), 1965–1977.  
462 URL <http://www.bioone.org/doi/full/10.1111/j.1558-5646.2008.00412.x?prevSearch=>
- 463 Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., Challenger, W., 2008. Bioinformatics applications  
464 note. *Bioinformatics* 24 (1), 129–131.
- 465 Iacus, S. M., 2008. *Simulation and Inference for Stochastic Differential Equations With R Examples*.  
466 Springer, New York.



- 467 Losos, J., December 1999. Uncertainty in the reconstruction of ancestral character states and limitations  
468 on the use of phylogenetic comparative methods. *Animal behaviour* 58 (6), 1319–1324.  
469 URL <http://www.ncbi.nlm.nih.gov/pubmed/10600155>
- 470 Losos, J. B., 1996. Perspectives on Community Ecology. *Ecology* 77 (5), 1344–1354.
- 471 Losos, J. B., 2009. *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiations of Anoles*.  
472 University of California Press, Berkeley, Los Angeles, London.
- 473 Martins, E. P., Hansen, T. F., 1997. Phylogenies and the Comparative Method: A General Approach  
474 to Incorporating Phylogenetic Information into the Analysis of Interspecific Data. *The American*  
475 *Naturalist* 149 (4), 646–667.
- 476 O’Meara, B. C., Ané, C., Sanderson, M. J., Wainwright, P. C., May 2006. Testing for different rates  
477 of continuous trait evolution using likelihood. *Evolution; international journal of organic evolution*  
478 60 (5), 922–33.  
479 URL <http://www.ncbi.nlm.nih.gov/pubmed/16817533>
- 480 Pagel, M., 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Compar-  
481 ative Analysis of Discrete Characters. *Society* (October 2009), 37–45.
- 482 Pagel, M., Meade, A., May 2006. Bayesian Analysis of Correlated Evolution of Discrete Characters by  
483 Reversible-Jump Markov Chain Monte Carlo. *The American naturalist* 167 (6).  
484 URL <http://www.ncbi.nlm.nih.gov/pubmed/16685633>
- 485 Pagel, M., Meade, A., Barker, D., 2004. Bayesian estimation of ancestral character states on phyloge-  
486 nies. *Systematic biology* 53 (5), 673–84.  
487 URL <http://www.ncbi.nlm.nih.gov/pubmed/15545248>
- 488 Paradis, E., 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20 (2),  
489 289–290.  
490 URL [http://www.bioinformatics.oupjournals.org/cgi/doi/10.1093/bioinformatics/  
491 btg412](http://www.bioinformatics.oupjournals.org/cgi/doi/10.1093/bioinformatics/btg412)
- 492 Schluter, D., Price, T., Mooers, A., Ludwig, D., 1997. Likelihood of ancestor states in adaptive radia-  
493 tion. *Evolution* 51 (6), 1699–1711.  
494 URL <http://www.jstor.org/stable/2410994>
- 495 Wainwright, P., 1988. Morphology and ecology: functional basis of feeding constraints in Caribbean  
496 labrid fishes. *Ecology* 69 (3), 635–645.  
497 URL [http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Morphology+  
498 and+Ecology:+Functional+Basis+of+Feeding+Constraints+in+Caribbean+Labrid+Fishes#0](http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Morphology+and+Ecology:+Functional+Basis+of+Feeding+Constraints+in+Caribbean+Labrid+Fishes#0)
- 499 Williams, E., 1969. The ecology of colonization as seen in the zoogeography of anoline lizards on small  
500 islands. *The quarterly review of biology* 44 (4), 345–389.  
501 URL <http://www.jstor.org/stable/2819224>