Inferring Adaptive Landscapes from Phylogenetic Trees: A Dissertation Proposal

Carl Boettiger[a,*]

[a] *Center for Population Biology, University of California, Davis, United States*

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.

Keywords: Evolution, Phylogenetics, Comparative Methods

1. Multiple niches can confound existing methods

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

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

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

---

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

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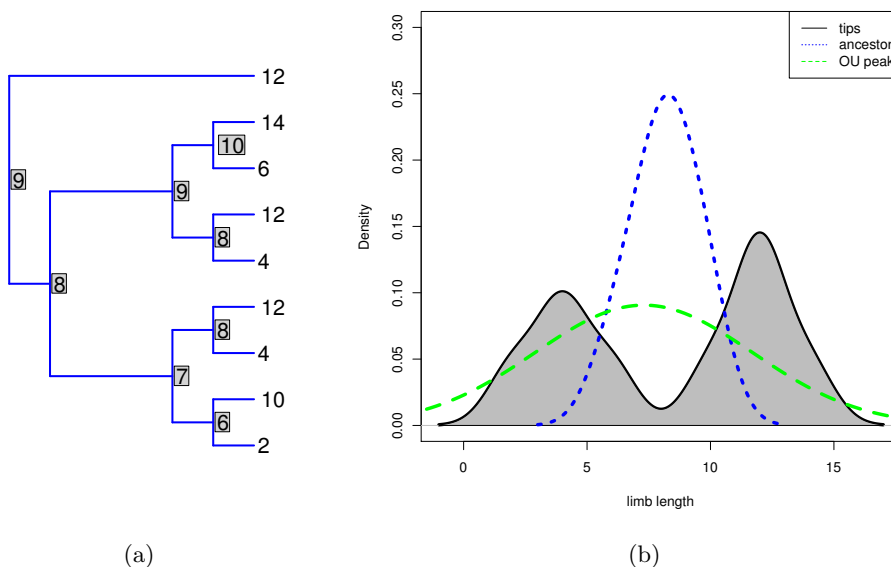

Figure 1: (a) Hypothetical Anolis hind limb lengths on a phylogenetic tree. Ancestral states inferred under Felsenstein's Brownian motion model of evolution are shown at the interior nodes. (b) Distribution of observed limb lengths at tips compared to the inferred ancestral states

The challenges introduced by phylogenies that span traits sampled from different niches requires a new approach to accurately account for the importance of phylogenetic relationships in using comparative data. In this, we follow an emerging pattern in the field of comparative methods. (1) Felsenstein (1985) first introduced BM as a model of trait evolution to enable existing applications of comparative methods, usually correlations between two traits or trait and environment, to account for the biological reality of phylogenetic relatedness. This approach was quickly adapted to answer novel questions such as estimating rates of diversification (Garland, 1992) and comparing these rates between clades (O'Meara et al., 2006; Collar et al., 2005). While the BM model provided a reasonable description of diversification, it could not capture the biological reality of stabilizing selection, and could consequently give spurious results when applied to traits that were strongly conserved over a phylogeny. (2) To address this, Hansen (1997) introduced the OU model as a description of adaptive trait evolution under stabilizing selection, which in turn introduced a new class of questions – identifying the strength
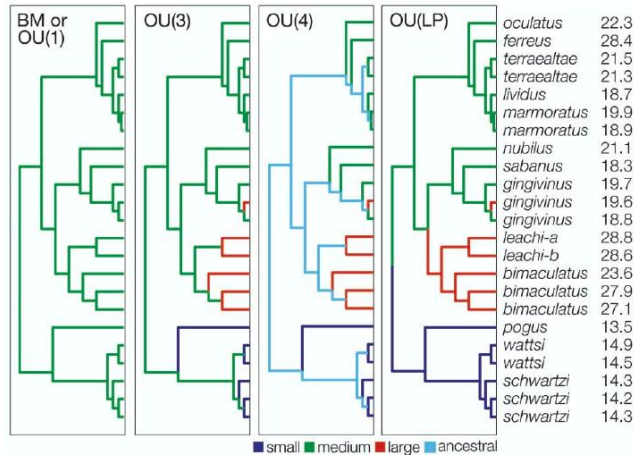
Figure 2: Possible paintings of a phylogenetic tree, taken from Butler and King (2004).

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

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

The introduction of OU model and the painting of multiple OU models onto a tree has allowed the field to enter the realm of model testing and model comparison. These comparisons are typically based on information criteria such as AIC, which penalize models that have more parameters by a certain factor. However, there is little reason to believe that AIC is fundamentally meaningful way to perform this discrimination. Using the data of Butler and King (2004) I fitted a Brownian motion model which I used to generate 1000 simulated datasets, then repeated the analysis on the simulated data. In 391 of the sets the best chosen model of Butler and King (2004), which uses three different OU models, still has the best AIC score, indicating the penalization for extra parameters is not sufficient. As can 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.

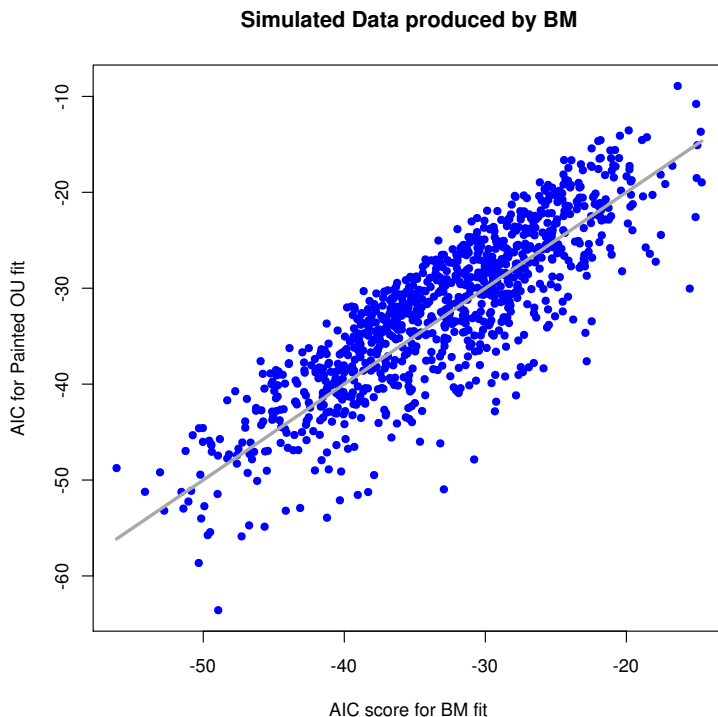**Simulated Data produced by BM**

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.

4

1. It matters which traits we include in the description. We are interested in describing functional traits. Measuring arbitrary traits and clustering along PCA is insufficient, as the functional trait may not be a linear combination of the morphology – indeed the "many-to-one" mapping that results from nonlinear interactions between morphology and performance is thought to be both pervasive and important in our understanding of evolution. Fish jaw morphology provides an excellent example, where performance can be well-characterized by suction index of a fish – an example of the "right" functional trait which is not a linear combination of basic morphological values.

2. We would like a robust quantitative way to define clusters along these trait axes. We should be able to determine how many clusters exist and quantify how well we can discriminate between peaks. Having identified interesting traits to test, there are two broad reasons why this clustering may not be immediately apparent – noise and phylogenetic inertia. Noise could come from many sources, from measurement error to environmental variation to the suite of other selective demands on the traits; all of which serve to blur out distinct clusters. Phylogenetic inertia may blur the distinction between groups if certain species have actively evolving in the direction of a particular cluster from the vicinity of another, and hence may not fall clearly into one group.

3. We would like to do so in a way which reflects the evolutionary process. This will allow us to better interpret the evolutionary significance of the parameters which define the clusters – for instance, strength of selection within groups and rates of transitions between groups.

4. We would like to utilize phylogenetic information to inform the clusters. If we know the phylogenetic tree connecting all species in the analysis, we may be able to use this information not only to correct for phylogenetic inertia but to penetrate the noise that may obscure clusters.

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

$$\mathrm{d}X = \alpha(\Theta - X)\mathrm{d}t + \sigma\mathrm{d}W_t \tag{1}$$

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

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

---

[1] 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
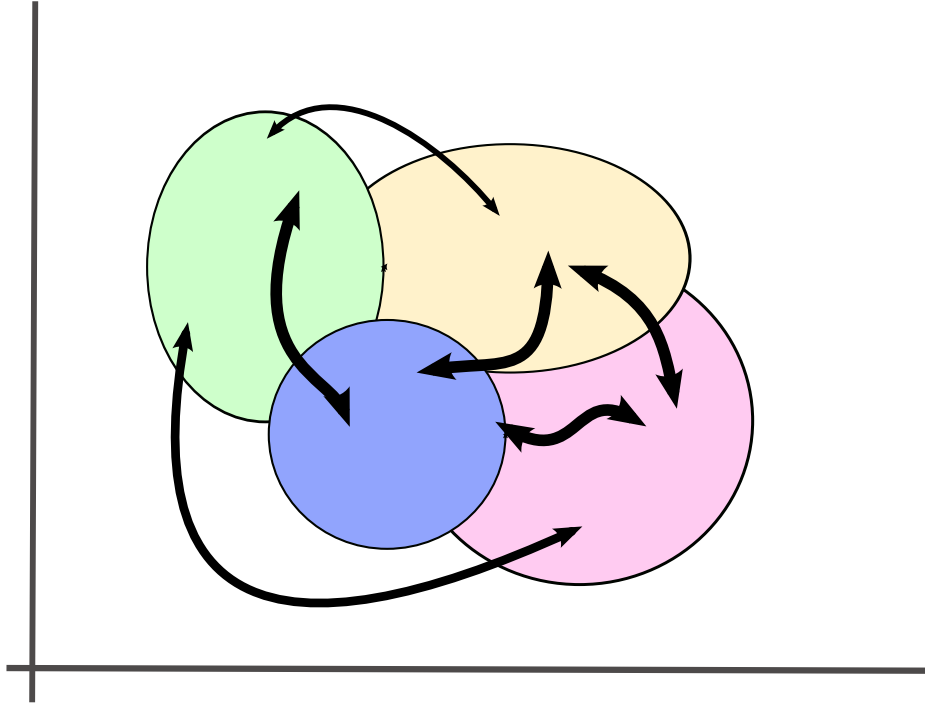
Figure 4: A conceptual model of four ecomorphs and transitions between them. Axes represent trait space, and ecomorphs differ in the location of their optima with respect to each trait and the strength of selection in each dimension. Rates of transitions may differ (as indicated by different weights) 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 α relative to diversification σ. Transitions could be captured as a Poisson process with transition rate matrix $\mathbb{Q}$.

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

3. Method

3.1. The right data

The data refers to a suite (vector) of observed traits across taxa their ultrametric tree. We will 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.

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

3.2. The Joint Probability

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

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

In a numerical implementation, this can be defined as a recursion over matrix multiplications. Assume the trait assumes values in an interval $\mathbb{I}$ which we discretize into m points. Then along each branch we can place the transition rate as an m by m matrix of the transition rates from any point in the space to any other during that time interval. As the values at the tips are known, we need only the column corresponding to transitions to the observed state. If the root node is specified (as a parameter of the model) then its branches are row vectors from the given trait to each point along the interval. The recursion starts at the root with this row vector and multiplies the row vector it receives from the previous node by the transition matrices of its left and right children. This proceeds up the tree until it reaches a tip, in which case the row vector multiplies the column vector of the tip to produce a scalar. The product of all these scalars is the joint probability.
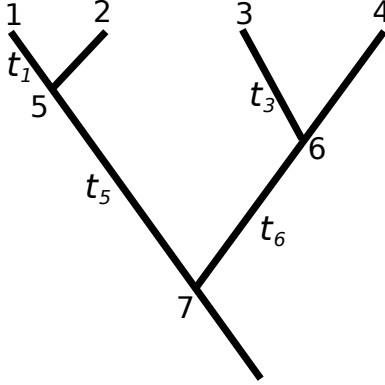
7

Figure 5: A simple phylogenetic tree.

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

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

$$(3)$$

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

4. Fitting the model

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

5. Example Systems

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

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

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

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

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

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

9

Labropsis australis   *Gomphosus varius*

(a)

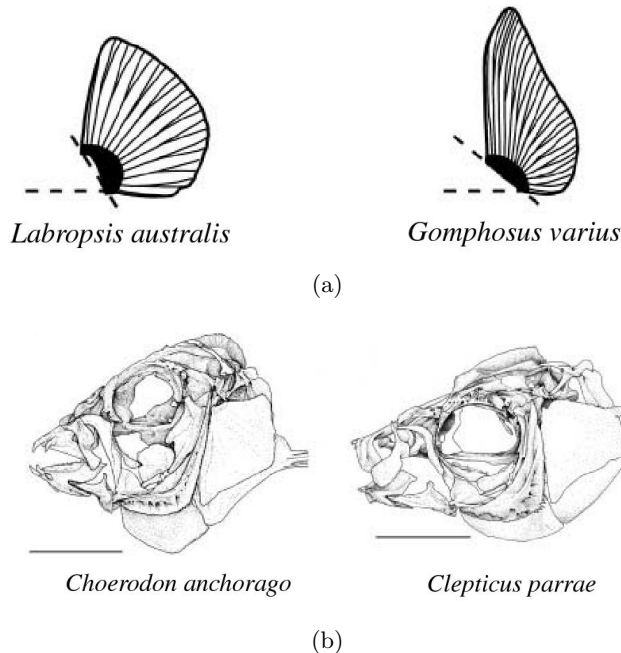*Choerodon anchorago*   *Clepticus parrae*

(b)

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).

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

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

6. Extensions

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

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

Some statistical validation should be part of any method. Particularly important here is the observation that any particular outcome is intrinsically unlikely, given the enormous number of possible outcomes of any model. Nevertheless, estimates of power are relatively straight forward, if computationally intense, to provide. Given a selected model with its maximum likelihood parameter values, one can create a set of simulated data sets and then run the method on this simulated data and compare the spread in parameter estimates from each of these runs. Creating simulated data sets is quite fast, though clearly the comparison will scale linearly with the number of simulated sets used. In the same fashion one can estimate the power of the model selection algorithm by asking for how many of the simulated data sets does the method select the same model as the best choice?
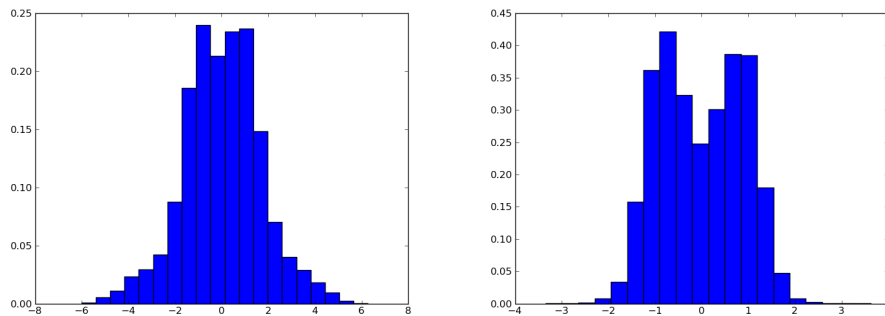
Figure 7: Variance due to random effects obscures the underlying two basins of attraction (niches) on the left. On the right, under reduced variance the shape of the adaptive landscape is more visible

6.2. Multi-dimensional traits

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

6.3. Discrete traits

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

6.4. Incorporating fossils

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

6.5. Changing Environments: Niche creation and loss

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

11

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

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

6.6. Bounded diversification and noise models

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

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

$$\mathrm{d}X = \beta \exp\left(\frac{-(X-\theta)^2}{2\sigma^2}\right)\mathrm{d}B_t, \tag{4}$$

a random walk where the size of the steps decreases away from the center of the range θ, that is, high diversification rates are found near the center of the trait distribution but decrease as one moves away.[2]

6.7. Ecologically realistic models

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

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

6.8. Scoring empirical performance landscapes

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

---

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

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

7. Significance and Broader Impacts

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

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

Appendix A. Model Library

Brownian Motion
$$\mathrm{d}X_t = \sigma \mathrm{d}B_t \tag{A.1}$$

Ornstein-Uhlenbeck
$$\mathrm{d}X_t = \alpha(\theta - X_t)\mathrm{d}t + \sigma \mathrm{d}B_t \tag{A.2}$$

N Gaussian Niches:
$$\mathrm{d}X_t = \left(\sum_i \alpha_i(\theta_i - X_t)e^{-\frac{(\theta_i - X_t)^2}{2\omega_i^2}} \right) \mathrm{d}t + \sigma \mathrm{d}B_t \tag{A.3}$$

Appendix B. Why the linear process is easy

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

$$\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(\mathrm{E}(X_i|X_z), \mathrm{E}(X_j|X_z)) \tag{B.1}$$

Because f is linear, it is straightforward to find the moments of the SDE. The conditional expectation $E(X_i|X_z)$ comes from solving the ODE with linear f and constant g, $\mathrm{d}\mathrm{E}(X_i) = \mathrm{E}(f(X_i))\mathrm{d}t$ with the initial condition $\mathrm{E}(X_i(0)) = X_z$. Similarly we can find the variance: for instance, for the OU process with initial condition $X(0) = X_a$ we have,

$$\mathrm{E}(X_t|X_a) = \alpha \int_0^t \theta(s)e^{-\alpha(t-s)}\mathrm{d}s + X_a e^{-\alpha t}, \tag{B.2}$$

which for constant θ is

$$\mathrm{E}(X_t|X_a) = \theta\left(1 - e^{-\alpha t}\right) + X_a e^{-\alpha t} \tag{B.3}$$

$$\mathrm{Var}(X_t|X_a) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}), \tag{B.4}$$

$$\mathrm{Cov}(X_i, X_j) = e^{-\alpha t_{ij}}\mathrm{Var}(X_a), \tag{B.5}$$

see Hansen (1997).

Bellwood, D. R., Wainwright, P. C., Fulton, C. J., Hoey, A. S., 2006. Functional versatility supports coral reef biodiversity. Proceedings of The Royal Society B 273 (1582), 101–7.
URL http://www.ncbi.nlm.nih.gov/pubmed/16519241

Butler, M. A., King, A. A., December 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. The American Naturalist 164 (6), 683–695.
URL http://www.journals.uchicago.edu/doi/abs/10.1086/426002

Collar, D. C., Near, T. J., Wainwright, P. C., August 2005. Comparative analysis of morphological diversity: does disparity accumulate at the same rate in two lineages of centrarchid fishes? Evolution 59 (8), 1783–94.
URL http://www.ncbi.nlm.nih.gov/pubmed/16329247

Collar, D. C., O'Meara, B. C., Wainwright, P. C., Near, T. J., 2009. Piscivory Limits Diversification of Feeding Morphology in Centrarchid Fishes. Evolution 63 (6), 15571573.
URL http://www.bioone.org/doi/full/10.1111/j.1558-5646.2009.00626.x?prevSearch=

Collar, D. C., Wainwright, P. C., Alfaro, M. E., 2008. Integrated diversification of locomotion and feeding in labrid fishes. Biology Letters 4 (1), 84.
URL http://intl-rsbl.royalsocietypublishing.org/content/4/1/84.full

Cunningham, C., September 1998. Reconstructing ancestral character states: a critical reappraisal. Trends in Ecology & Evolution 13 (9), 361–366.
URL http://linkinghub.elsevier.com/retrieve/pii/S0169534798013822

Felsenstein, J., 1985. Phylogenies and the comparative method. The American Naturalist 125, 1–15.
URL http://www.journals.uchicago.edu/doi/abs/10.1086/284325

Freckleton, R. P., Harvey, P. H., 2006. Detecting Non-Brownian Trait Evolution in Adaptive Radiations. PLoS Biology 4 (11), 2104–2111.

Garland, T. J., December 1992. Rate Tests for Phenotypic Evolution Using Phylogenetically Independent Contrasts. The American Naturalist 140 (3), 509 – 519.

Gavrilets, S., Losos, J. B., 2009. Adaptive radiation: contrasting theory with data. Science 323, 732–737.
URL http://science.samxxzy.ns02.info/cgi/content/abstract/323/5915/732

Gavrilets, S., Vose, A., December 2005. Dynamic patterns of adaptive radiation. Proceedings of the National Academy of Sciences 102 (50), 18040–5.
URL http://www.ncbi.nlm.nih.gov/pubmed/16330783

Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82 (4), 711.
URL http://biomet.oxfordjournals.org/cgi/content/abstract/82/4/711

Green, P., 2003. Trans-dimensional markov chain monte carlo. Vol. 27. Oxford University Press, p. 179198.
URL http://books.google.com/books?hl=en&lr=&id=NecQlPWrN3AC&oi=fnd&pg=PA179&dq=Trans-dimensional+Markov+chain+Monte+Carlo&ots=wOnwaRZKQd&sig=O8sciWOMmaUqRl6eEzDltzTgzLo

Hansen, T. F., 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51 (5), 13411351.
URL http://www.jstor.org/stable/2411186

Hansen, T. F., Martins, E. P., 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50 (4), 14041417.
URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Translating+Between+Microevolutionary+Process+and+Macroevolutionary+Patterns#0

Hansen, T. F., Pienaar, J., Orzack, S., 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62 (8), 1965–1977.
URL http://www.bioone.org/doi/full/10.1111/j.1558-5646.2008.00412.x?prevSearch=

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., Challenger, W., 2008. Bioinformatics applications note. Bioinformatics 24 (1), 129–131.

Iacus, S. M., 2008. Simulation and Inference for Stochastic Differential Equations With R Examples. Springer, New York.

Losos, J., December 1999. Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. Animal behaviour 58 (6), 1319–1324.
URL http://www.ncbi.nlm.nih.gov/pubmed/10600155

Losos, J. B., 1996. Perspectives on Community Ecology. Ecology 77 (5), 1344–1354.

Losos, J. B., 2009. Lizards in an Evolutionary Tree: Ecology and Adaptive Radiations of Anoles. University of California Press, Berkeley, Los Angeles, London.

Martins, E. P., Hansen, T. F., 1997. Phylogenies and the Comparative Method: A General Approach to Incorporating Phylogenetic Information into the Anlysis of Interspecific Data. The American Naturalist 149 (4), 646–667.

O'Meara, B. C., Ané, C., Sanderson, M. J., Wainwright, P. C., May 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60 (5), 922–33.
URL http://www.ncbi.nlm.nih.gov/pubmed/16817533

Pagel, M., 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. Proceedings of The Royal Society B 255, 37–45.

Pagel, M., Meade, A., May 2006. Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. The American Naturalist 167 (6).
URL http://www.ncbi.nlm.nih.gov/pubmed/16685633

Pagel, M., Meade, A., Barker, D., 2004. Bayesian estimation of ancestral character states on phylogenies. Systematic biology 53 (5), 673–84.
URL http://www.ncbi.nlm.nih.gov/pubmed/15545248

Paradis, E., 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20 (2), 289–290.
URL http://www.bioinformatics.oupjournals.org/cgi/doi/10.1093/bioinformatics/btg412

Schluter, D., Price, T., Mooers, A., Ludwig, D., 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51 (6), 16991711.
URL http://www.jstor.org/stable/2410994

Wainwright, P. C., 1988. Morphology and ecology: functional basis of feeding constraints in Caribbean labrid fishes. Ecology 69 (3), 635645.
URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Morphology+and+Ecology:+Functional+Basis+of+Feeding+Constraints+in+Caribbean+Labrid+Fishes#0

Wainwright, P. C., Bellwood, D. R., 2001. Locomotion in labrid fishes: implications for habitat use and cross-shelf biogeography on the Great Barrier Reef. Coral Reefs 20, 139–150.

Williams, E., 1969. The ecology of colonization as seen in the zoogeography of anoline lizards on small islands. The quarterly review of biology 44 (4), 345 389.
URL http://www.jstor.org/stable/2819224