

# PhD Literature Review MS

## **Contents**

# 1 Introduction

Evolution is heritable change through time. Over long timescales, the discrete objects subject to change are species. Phylogenies are trees showing species, speciation events and their arrangement in time. Since the advent of evolutionary theory, species have been compared with each other to suggest adaptive explanations for their trait values. Similar trait values, however, can also be due to shared ancestry; hence, we infer jointly the phylogenetic and ecological effects on trait values. This is aided by the wealth of reliable phylogeny data now obtainable from DNA sequencing.

To infer evolutionary processes from trait data, we require models of these processes that predict trait data. These models have two forms: they can predict trait evolution along a branch of a phylogeny, or they can predict the times at which branches will bifurcate.

The evolutionary models are outlined here, along with models which associate discrete amounts of evolutionary change with speciation events. The biological expectations and evidence for these models are discussed, and future developments are briefly considered.

## 2 Models of continuous evolution

### 2.1 Data

Models of continuous change on a phylogeny rely on three pieces of data:

- Tree topology;
- Branch lengths;
- Trait values of tree tips.

These models give instantaneous rates of evolution as a function of position on the phylogeny branch. Then, the difference between two species is predicted from the length of branches separating them. The tendency for phylogenetically close species to be phenotypically similar is referred to as ‘phylogenetic signal’ (?). Phylogenetic niche conservatism (PNC) is a related concept, with no universal definition (?). Loosely, it is phylogenetic signal where the trait is the species’ niche. Phylogenetic signal is often assumed to be particularly strong for such traits, but this assumption may be frequently unsatisfied (?).

The best data for comparative analyses consists of many closely related pairs of species that differ in trait value, with each pair well separated phylogenetically from the others (?). Ancestral trait values can be very useful in fixing non-tip phylogeny nodes, but are rare and generally not incorporated into phylogenetic comparative analyses (?).

## 2.2 Evolution on branches

The need to account for phylogeny in comparative analyses was first made clear by ?. He used a model where the change in trait value in a short period of time is drawn from a normal distribution: this is Brownian Motion (BM). Other models of trait evolution along a branch are derived from the BM model by adding parameters. Here, the various models are listed in time-differential form in Table 1.

The BM model has a trait value  $X$  evolving at random, at a rate  $\sigma$ :

$$dX(t) = \sigma dW(t) \quad (1)$$

where  $W(t)$  is the integral of the continuous white noise function, such that  $\Delta W_t \sim N(0, \Delta t)$ . If used to predict trait values at a time  $T$ , the BM model has two free parameters: the evolutionary rate  $\sigma$  and the root trait value  $X(0)$ .

The parameter  $\lambda$  (??) measures the goodness of fit of trait data to those expected under a given BM model. A value of 1 indicates a perfect fit to the BM model;  $\lambda < 1$  indicates less phylogenetic signal and  $\lambda > 1$  indicates more signal than expected.  $\lambda$  can be represented as a transformation of the tree into one with internal branches scaled by the factor  $\lambda$ .

We know that rates of evolution vary. How do we add plausible patterns of rate-change to the BM model? Looking at equation 1, we seem to have two options: we can elevate the factor  $\sigma$  to the status of a function, or we can add a separate, deterministic term to the equation. Time-dependency of  $\sigma$  could be linear ( $\sigma \rightarrow (t_{max} - t)\sigma$ ), polynomial ( $\sigma \rightarrow \kappa t^{\kappa-1}\sigma$ ), or exponential ( $\sigma \rightarrow g^{-t}\sigma$ ).

**Figure 1.** An example of evolution under the BM (black) and OU (red) models.

Polynomial time dependence  $\kappa$  is first seen in ?. Exponential models, meanwhile, are termed Accelerating/Decelerating (ACDC) models (?). Each of these models adds one free parameter to the BM model. One more alternative exists and has been implemented: the evolutionary rate  $\sigma$  may be a step function of time; these steps (i.e. discrete rate changes) may represent sudden environmental changes, or transitions into or out of ecological niches (??). If some extrinsic event is known about, then the positions of the corresponding steps can be built into the model before the model is used. Alternatively, maximum likelihood (ML) positions and sizes of discrete rate changes can be estimated from the data (??).

The second way to modify the BM model of equation 1 is to add an additional term. In the OU model (?) the trait  $X$  is drawn towards a central value ('primary optimum')  $\psi$  with a strength proportional to its distance from the optimum:

$$dX(t) = -\alpha(X(t) - \psi)dt + \sigma dW(t) \quad (2)$$

For this reason, the OU model is sometimes called the 'single stationary peak' (SSP) model. It is perhaps easier to imagine this as an ecological process of avoiding extreme

**Figure 2.** Variance among tip trait values under different models of evolution. Branch lengths in units of time. The black line is the BM model; the red line is  $\kappa < 1$ ; the blue line the OU model, with a different optimum for each branch; the green line is a NF model.

trait values: as a species gets further from the clade mean, it becomes more likely to evolve towards the clade mean. (Being repelled by two extreme values is of course equivalent to being attracted to one central value).

The OU model can be extended by adding discrete optima shifts based on prior knowledge of extrinsic events, or with methods to estimate ML positions of optima shifts (??). Additionally, the optima may themselves evolve according to a BM or OU model ??.

### 2.3 Evolution on phylogenies

Most of the BM-like models of evolution can be applied separately to each branch, but we can also make evolutionary rates dependent on position in the whole phylogeny. Firstly, we might test for a phylogeny-wide change in the rate of evolution, analogous to the within-branch parameter  $\kappa$  (?). This parameter takes the same form as  $\kappa$ , and is referred to as  $\delta$  (?) (See Table 1). We can also permit rate parameters to be branch-specific. Having an independent  $\sigma$  for each branch is possible (?), but will tend to lead to too many free parameters (?). ML estimation of a limited number of discrete rate changes, however, can be applied clade-wise as well as time-wise, so that a change appears on just one branch in a phylogeny, but is inherited by subsequent ‘offspring’.

### 2.4 Stretching branches

If phylogeny branch lengths represent time, then the evolutionary models in the previous section allow us to calculate the amount of change per branch - not necessarily proportional to time. The evolutionary models can also be viewed as maps from time-scaled trees into change-scaled trees. The BM model is then the identity operator. Some of the models were originally presented as such a transformation. An example tree transformed under each model is shown in Figure 3.

**Figure 3.** Change-scaled trees. The BM tree is identical to the time-scaled tree, and different evolutionary models are different transformations into other trees. Kappa and delta are between 0 and 1.

### 2.5 Multiple traits

The BM and BM-like models generalise immediately to multiple traits: the trait value  $X$  becomes a vector of trait values  $\vec{X}$ , and the rate parameter  $\sigma$  becomes a covariance

matrix  $\Sigma$ . Nonzero nondiagonal elements of  $\Sigma$  indicate correlation between traits.

One of the main uses for phylogenetic methods is to measure correlations between traits while controlling for the phylogeny. This is discussed in Section 4, but it may be useful to note a potential source of confusion: many methods seek to separate phylogenetic and ecological effects. However, ecology may be heritable. If two species are labile but have phylogenetic signal due to shared inherited ecology, then they are independent data points with respect to the details of that ecology, but non-independent with respect to general ecological principles.

OU model variants have been developed to account for coadaptation between traits, with trait optima either fixed or evolving by the BM model. With multiple traits,  $\alpha$  becomes a matrix, and off-diagonal elements can permit one trait’s optimum to change according to another trait’s value, even if the second trait is simply evolving according to BM (?).

## 3 Models of speciation-based evolution

### 3.1 Evolution on phylogenies

Branch-evolution parameters such as  $\kappa$  can be used as a test for speciation evolution. An alternative is simply to postulate a ‘lump’ of trait change at each speciation event, with the lump drawn from  $N(0, \sigma_c)$  (??). If the speciation rate is  $\xi$ , then we can measure the degree to which rapid evolution is linked to speciation events with the value of

$$\frac{\xi \sigma_c^2}{\sigma_{total}^2} \quad (3)$$

This model requires extinction rate estimates. Methods to generate extinction rate estimates exist (??), but it is not clear how to get these without first making assumptions about the evolutionary model that we are hoping to test.

If the amount of Brownian Motion is set to zero, then the ? model reduces to random amounts of trait change at speciation events and none in between speciation events. This corresponds to a pure ‘punctuated equilibrium’ model of evolution.

Niche-filling (NF) models (???) also have purely speciation modes of evolution, but have a few differences from the Ingram model. In niche-filling models, species’ trait values are constant in time, and the position of each niche in niche-space is constant in time. New niches are filled by new species, branching from whichever species is closest in niche-space. ? distinguish between NF models where every species is equally likely to speciate and form a new and filled niche, and NF models where niches appear at a constant rate, randomly positioned in niche-space. The two are equivalent when all the species have evenly distributed trait values, but have different predictions otherwise; for example, a species particularly distant from all others in trait-space will be more likely to speciate under the randomised NF model.

In the Ingram model speciation is a splitting process, where each of the two species shifts to a new trait value. Hence, in NF models speciation is a branching process, where only the ‘new’ species moves to a new trait value. This leads to a difficulty for NF models, however: if a lineage splits and the ‘original’ lineage subsequently goes extinct, then a phylogeny built from extant species will have unexplained mid-branch evolutionary change (assertion).

A further model combines phylogenetic signal from a BM process with a postulate of trait variance increasing linearly with spatial separation between the species (?). The relative importance of the two terms of the expression for the variance then measures the importance of physical locale in the evolution of the trait.

In niche-filling models, species respond very quickly to environmental changes (via speciation), but phylogenetic signal exists and persists, because new niches are filled from nearby niches. Hence, sibling species resemble one another. The BM model achieved this signal by the contrary postulate that evolution is purely random and unconstrained.

## 3.2 Speciation and phylogeny shape

BM-like models require a phylogeny as input data, and provide no insights on the question of when and why speciation should occur. If  $b(t)$  is the birth rate and  $d(t)$  is the death rate, then a ‘Yule process’ has  $d=0$  and  $b$  constant so that the number of species increases linearly with time. The  $\gamma$  statistic tests for deviations from the Yule model; it is a measure of acceleration in the rate of speciation, with standard normal distribution under the Yule model (?).

A reduction in speciation rate through time is consistent with NF models that accommodate a limited number of niches. However, this pattern can be seen even in models which do not have such ecological effects; therefore, it is important to compare model predictions and not just test a null model of constant speciation rate (?). One solution is to look for co-occurring slowdowns in speciation rate and trait evolution rate (?). However, if trait values are linked to speciation rate, evolutionary rate estimates will be biased (?).

# 4 Comparing models

## 4.1 Relating models to data

Given a tree and an evolutionary model, we want to know the probability of obtaining the observed phenotypic tip data. The commonest method is independent contrasts, developed by ??. The difference in trait value between sibling species depends only on the branch length (in units of expected evolutionary change) separating them, and is therefore independent of shared evolutionary history. Since trees bifurcate, each node has exactly one sibling, so there are  $n-1$  independent contrasts in a tree with  $n$  tips. The uncertainty in non-tip node trait values is accommodated by lengthening their branch by

**Table 1.** Models and parameters for evolution on a known phylogeny. Multi-trait parameter counts are upper limits assuming free covariance between all traits.

Model	Equation	parameters	parameter count
BM	$dX(t) = \sigma dW(t)$ $var(X) = T\sigma^2$	root value, $\sigma$ =rate	2
$\kappa$	$dX(t) = \kappa\sigma t^{\kappa-1}dW(t)$ $var(X) = \Sigma_i T_i \sigma$	T is time from common ancestor $\kappa > 0$ , t from branch start i labels each branch, and $T_i$ is its length	3
$\delta$	$dX(t) = \delta\sigma t^{\delta-1}dW(t)$ $var(X) = \sigma^2(T^\delta - T_0^\delta)$	$\delta > 0$ , t from tree root T is time over tree, $T_0$ is time of MRCA	3
Step	$dX(t) = \sigma(t, clade)dW(t)$	$\sigma$ step fn with m steps	1+2m
ACDC	$dX(t) = \sigma g^{-t}dW(t)$	$g > 0$	3
OU	$dX(t) = -\alpha(X(t) - \psi)dt + \sigma dW(t)$ $var(X) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha T})$	$\psi$ is optimum, strength $\alpha$	4
stepOU	$dX(t) = -\alpha(X(t) - \psi_i)dt + \sigma dW(t)$	i is number fitted optimum shifts	4
Ingram2010	BM + $N(0, \sigma_c^2)$ per speciation		3

an amount  $v_i v_j / (V_i + v_j)$ , where  $v_i, v_j$  are the branch lengths of the species descended from that node. Methods exist to use trees with unresolved nodes (?).

The rate  $\sigma^2$  of Brownian evolution can be estimated with a generalised least squares (GLS) method (?). This method, equivalent to the independent contrasts method, uses regression such that each tip has a trait value

$$X_i = a + \sigma^2 \Sigma v_j + e_i, \quad (4)$$

where  $V_j$  is the total branch length separating tip i from tip j, and  $e_i$  is a residual. The evolutionary rate can then be estimated as

$$\sigma^2 = \frac{1}{n-1} X^T \mathbf{V}^{-1} X \quad (5)$$

where  $X$  is a vector of the tip trait values, and  $V$  is the covariance matrix, i.e. a matrix of the branch lengths (for non-BM models, the transformed branch lengths) shared by each pair of tips. Non-phylogenetic methods are the subset of GLS methods with  $V$  diagonal.

The autocorrelation method (?), like GLS, uses regression, partitioning the between-species trait variance into heritable (phylogenetic) and specific components. Then, covariance between traits of specific, but not heritable, components is evidence for coadapted traits. A covariance matrix is used, but not derived. The autocorrelation method generally performs less well than the independent contrasts method (?).

The animal model, or mixed model, used in quantitative genetics, can be adapted for comparative phylogenetic analysis (?). A species' trait value is multiply regressed on phylogenetic effect and the values of other traits, with additional residual terms. Efficient calculational methods exist, and within-species variation is readily included (?). The latter point is important: two identical populations imperfectly sampled will look different, causing independent contrasts to be overestimated. Restricted ML techniques now exist to correct for this within the independent contrasts method (??). This will be particularly important when comparing the BM and OU models, since the variation around a 'primary optimum' present in the OU model will resemble this bias.

The independent contrasts method removes phylogenetic effects without estimating them. This makes it computationally faster, and generally better performing when assumptions are broken. Methods which use the covariance of each tip with every other tip apply the evolutionary model to every point in the tree, not just the section containing the taxa under comparison. However, if the phylogenetic component of a species' trait is what we want to know, then regression methods are more appropriate. As evolutionary models become more complicated, and rates of evolution are modelled as functions of trait values, it seems likely that computer simulation will become preferred to the above methods. The main limitation of simulation is simply computational time (?).

When studying correlations between traits, it is not obvious that phylogenetic methods are superior to simple analyses of raw data. This is because phylogenetic methods, including independent contrasts, make assumptions about the evolutionary process. If



these assumptions are false, there are conditions under which analyses of raw data can be more accurate than those with erroneous phylogenetic corrections (??). This is one reason for using tests for phylogenetic signal (??). However, phylogenetic methods in general, and independent contrasts in particular, are usually well supported and robust to perturbation away from their assumptions (???).

One difficulty with most approaches is that they assume perfect knowledge of the phylogeny. Methods exist to account for unresolved nodes [ref]. However, this does not make use of the likelihood data which is usually generated by sequence-data phylogeny building. A recent study (?) has demonstrated the possibility of using a posterior distribution of phylogenies to account for phylogenetic uncertainty.

## 4.2 Model likelihoods

The likelihood of hypothesis  $H$  given data  $D$  is  $L(H|D) = \frac{P(D|H)P(H)}{P(D)}$ . A ratio of hypothesis likelihoods is then  $\frac{L(H_1|D)}{L(H_0|D)} = \frac{P(D|H_1)P(H_1)}{P(D|H_0)P(H_0)}$ . With no prior expectations of model likelihoods, the ratio becomes  $\frac{P(D|H_1)}{P(D|H_0)}$ . Models with more free (fitted) parameters should fit the data better. To avoid overparameterisation, we therefore have to require a 'significant' improvement in fit from the more complex model than the model. There are various approaches to determining this significance, including likelihood ratio tests (LRT), the Akaike information criterion (AIC) and Bayesian methods.

Using LRTs to compare two models results in a test statistic which is the log of the ratio of their likelihoods. When  $H_0$  is a special case of  $H_1$ , so that the models are 'nested', then the test statistic forms (0.5 times) a  $\chi^2$  distribution. This distribution, however, also assumes large samples, that one of the models is true, and that parameters are normally distributed; these assumptions may sometimes be significantly violated by phylogenetic methods (?). To avoid these difficulties, we can take the parameter MLEs from the null model, and simulate new datasets from those parameters ('parametric bootstrapping', a Monte Carlo technique). For each dataset, new MLEs are generated, and the log-likelihood ratio calculated. The distribution of LRTs then allows us empirically to map LRT values to p-values.

The AIC is a number assigned to each model: the difference between the maximised log-likelihood and the number of free parameters  $K$ :

$$AIC = -2l + 2K. \quad (6)$$

'Akaike weights' then represent relative likelihoods of models:

$$w_i = \frac{e^{-\Delta_i}}{\sum_{r=1}^R e^{-\Delta_r}} \cdot \Delta_i = AIC_i - \min(AIC) \quad (7)$$

Like the LRT, the AIC assumes a large sample size with parameters that are multivariate normal (?). However, the AIC has a key advantage over LRTs in that non-nested models can be tested without the need for simulation and bootstrapping.

When likelihoods themselves are difficult to calculate, we can estimate them by using the model to generate new simulated datasets, and comparing this distribution of datasets with the observed data. We can then choose to use the likelihood for the best-fit model parameters, as in the LRT, or to integrate over all model parameters according to a prior distribution of parameter values, chosen before fitting the model. One implementation of the latter approach is approximate Bayesian computation (ABC). In ABC, a set of parameter values is sampled from the prior distribution, and some data  $\hat{D}$  is simulated. For observed data  $D$  and tolerance  $\epsilon$ , we accept  $\hat{D}$  if

$$\rho(\hat{D}, D) \leq \epsilon, \quad (8)$$

where  $\rho$  is the discrepancy, or distance in solution space, between  $\hat{D}$  and  $D$ . The set of parameter values which produce accepted instances of  $\hat{D}$  are then taken to be a sample from the posterior distribution of parameter values. To compare models, each model's likelihood is taken to be proportional to the fraction of simulations accepted. Then we can use the Bayes factor:

$$K = \frac{P(D|M_1)}{P(D|M_2)} = \frac{\int P(\theta_1|M_1)P(D|\theta_1, M_1)d\theta_1}{\int P(\theta_2|M_2)P(D|\theta_2, M_2)d\theta_2} \quad (9)$$

where  $\theta$  is the set of parameter values.

Using fitted-model likelihood ratios, and integrated-model likelihood ratios represent two different measures of model usefulness, and it is probably advisable to calculate both and compare in order to learn more about the truth. By generating data under MLE model parameters, we can also visualise the distribution of modelled data alongside the observed data, to gain a idea of the model's adequacy in describing real data.

### 4.3 Empirical tests

A review of comparative studies found that  $\lambda$  is typically high, consistent with strong phylogenetic signal (?). This suggests limited applicability of the OU model, which predicts decay of signal. The presence of phylogenetic signal suggests that a BM or NF model will typically be best, but does not automatically distinguish between them (?). This question depends on evolutionary gradualism; for example, a recent study shows that two-thirds of variation in body mass is speciational (?). It would be interesting to see how  $\lambda$  varies with phylogeny shape across many comparative studies. However, it is important to make phylogenetic signal estimates jointly with ecological models, not prior to fitting the ecological models (?). An additional source of phylogenetic signal can be spatial effects: if closely related species also tend to be geographically closer then they may share adaptations to that local environment (?).

BM models have successfully been rejected in favour of NF models for data on warbler birds using two tests (?). Firstly, tests for correlation between independent contrasts (?).

and phylogenetic positions of the contrasted species reveal links between divergence rates and phenotype (i.e. position in niche-space). Secondly, testing for an overall slowdown of evolution across the phylogeny can reveal constraints arising from the available niches getting ‘filled up’. Other studies also find speciation rate slowing with time, consistent with a limited number of niches being filled, but find that phenotypic evolution does not share this slowdown (?). This pattern might be consistent with a NF model with niche positions evolving randomly in trait space. Alternatively, it could be that the niche is defined by a complex combination of traits, such that that combination is conserved while individual trait values are not (?).

Ecological release (the removal of selective constraints) is sometimes linked to adaptive radiations, but frequently is not (?). The chances of adaptive radiations may depend on fluctuations in population size and density, in turn dependent on fluctuations in selection strength and direction (??). Since long-term stasis can arise from short-term fluctuations, evolution can be the result not of environmental change but of environmental stability (?). ? conclude that many factors affect the link between ecological release and speciation, and that further study should assess the commonness of these factors, and their ability to reinforce or cancel each other. They recommend that population-genetic parameters such as effective population size and trait variance be included in models of long-term evolution.

## 5 Software

Software exists to visualise, simulate and fit evolutionary models to phylogenies and tip data. Most of this software exists as packages for the R platform (?). Trees can be stored as Newick files, with all tips extant and nested branch lengths. Nexus files can contain various data including Newick trees. The ape format (?) is an alternative which also allows nodes to be labelled, and arbitrary tip dates to be set, by labelling each branch by its parent and offspring species and specifying branch lengths explicitly. Phylo4d (?) adds to the ape format the ability to associate sequence or trait data with the tips.

GEIGER (?) is used to generate simulated phylogenies and trait data. It can randomly ‘prune’ clades to mimic incomplete sampling. The trees are created with birth-death models, and the tip data are generated under the BM model, with discrete and continuous traits. Multiple continuous traits can be simulated given a covariance matrix. GEIGER can perform AIC tests for significantly nonzero rate-change parameters.

The caper package ? allows model fitting to trees and tip data using the independent contrasts and the GLS methods. It uses ape data, and requires the ape package. The transformation parameters  $\delta, \kappa, \lambda$  can be estimated and tested with ANOVA or AIC. Trees can be simulated with birth-death models, and tip data can be simulated with a BM model.

The OUCH package (?) fits an OU model to tree and tip data, with  $\alpha$  and  $\sigma$  as

free parameters. Clades can be chosen to have independent estimations. Multivariate estimates can be made with a symmetric  $\alpha$ -matrix and a lower-diagonal  $\sigma$ -matrix.

MOTMOT (?) is a package used to fit a specified number of discrete rate changes in evolutionary rate, and to compare likelihood with the pure BM model. Both the size and position of the changes are estimated. This contrasts with the OUCH package, where clades must be chosen for independent parameter fitting beforehand. MOTMOT can also be used for AIC tests of nonzero continuous rate change parameters, and for ?'s  $\psi$  parameter. MOTMOT calculates the likelihood in closed-form solution for a rate-change at each tree node. Another package, phytools (??) fits discrete rate changes with an MCMC approach, permitting mid-branch rate changes.

## 6 Conclusion

There is an emerging appreciation of the importance of addressing interactions between evolutionary and ecological processes (?). Predictors of *what* is adaptive for a species will depend on the world in which it lives - in other words, its ecology. The BM model is therefore an ecologically neutral model. To test ecological models, we look for deviations from the BM model. Defining ecology broadly as the interactions of a species with its environment (including other species), this is true by definition.

Evolutionary models generally try to fit one or a few evolutionary processes to a relatively large phylogeny. However, we know that evolution is associated both with branch length and with speciation (?). Recent studies demonstrate evolution that is potentially rapid but actually very constrained (??). It seems likely that this should result in wide variability in rate and pattern of evolution between lineages. The unsorted tip data may be then well approximated by a BM model, and the BM model is an accepted null model for such ecological tests (?). But, if we can model some ecological processes, lineage by lineage, a more predictive model may be attainable. Population size can determine whether constraints stop or merely slow evolution (?), and such 'microevolutionary' parameters which have been recommended for inclusion in evolutionary models (?). It makes sense to extend existing NF models, which track lineages individually, but as a function of randomly occurring niches.

In addition to the work which has been done on trait evolution, there is a separate body of literature on speciation and extinction rates, including methods to model these rates as functions of trait values (????), and to link these rates to ecological community assembly

The key emergent foci for future work include:

- How can models of trait evolution and models of speciation and extinction rates best be integrated?

- Can the results of existing likelihood techniques be easily replicated with simulation methods such as ABC, and what is the speed penalty?
- What effects do population demographics have on trait evolution? Can these effects be inferred from phylogenetic comparative analyses using demographic inferences from sequence data?
- How can niche-filling models be parameterised to include ecological processes such as competition, community construction and ecological release?