Evolutionary Inferences from Phylogenies: A Review of Methods

Brian C. O'Meara

Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996; email: bomeara@utk.edu, **http://www.brianomeara.info**

Annu. Rev. Ecol. Evol. Syst. 2012. 43:267–85

First published online as a Review in Advance on September 4, 2012

The *Annual Review of Ecology, Evolution, and Systematics* is online at ecolsys.annualreviews.org

This article's doi: 10.1146/annurev-ecolsys-110411-160331

1543-592X/12/1201-0267$20.00

Keywords

methods, continuous-time Markov Chain, multivariate normal, birth-death, tree stretching

Abstract

There are many methods for making evolutionary inferences from phylogenetic trees. Many of these can be divided into three main classes of models: continuous-time Markov chain models with finite state space (CTMC-FSS), multivariate normal models, and birth-death models. Numerous approaches are just restrictions of more general models to focus on particular questions or kinds of data. Methods can be further modified with the addition of tree-stretching algorithms. The recent realization of the effect of correlated trait evolution with diversification rates represents an important advance that is slowly revolutionizing the field. Increased attention to model adequacy may lead to future methodological improvements.

INTRODUCTION

A phylogeny is an inferred history of species splitting through time. It may just show the relationships of the species, but it may also have branch lengths (in units of time, number of generations, or amount of change) and, more rarely, widths (population size through time). As species evolve up a phylogeny, various processes lead to changes in trait values within populations. For example, the optimal body size may change as climate warms, as competitors invade a habitat, or as new resources become available, so the average body size of individuals in a species can change through time as the population evolves toward the moving optimum.

Information about a phylogeny and trait values at the tips of the tree can be combined with an inference method to learn about evolutionary processes. One frequently applied example of this is methods for inferring character states for ancestral species. These methods have been used to reconstruct what calls extinct frogs may have made and to evaluate the reaction of extant frogs to these reconstructions to examine whether frogs evolve to sound unlike their neighbors (Ryan & Rand 1995). Similarly, the history of changes can be inferred from a tree, as was done recently with swine flu viruses to examine how certain strains lead to epidemics (Vijaykrishna et al. 2011). Methods that investigate the correlation of gene presence over evolutionary time can be used to predict functional links between genes better than methods that ignore a tree (Barker & Pagel 2005).

Some questions that can be addressed using this sort of approach include what trait the ancestor to a particular group of species had, how the possession of trait W influences the evolution of trait X, whether trait Y leads to a higher diversification rate, which group is evolving phenotypically the fastest, which group is diversifying the slowest, whether the rate of evolution has decreased through time, whether two particular clades are evolving toward different evolutionary optima, what the relative influences of past and present optima are on the observed value of a trait, whether trait Z is rarely observed because species with this trait go extinct rapidly, whether population size has changed through time, whether there has been gene flow between particular putative species, when a species dispersed between particular continents, and many more. However, though there is a dizzying array of methods for using phylogenies to make evolutionary inferences, certain themes keep arising. Rather than list all the methods, this review describes the connections between many methods. For example, the same basic model can be used for estimating whether an ancestral species had wings (ancestral state reconstruction), whether evolution of woody stems leads to evolution of bat-dispersed fruits (character correlation), and which tree best fits the data (DNA evolution). Many of the most widely used methods rely upon one of the three following classes of processes (which could themselves be grouped at deeper levels): continuous-time Markov chains with a finite state space (CTMC-FSS), multivariate normal distributions, and birth-death branching processes. Some important methods, however, do not fall neatly into these three bins. A set of approaches we can call tree stretching, the transformation of some or all branches of a tree, is a frequently employed approach used to build new methods from these common classes. New approaches jointly examine character transitions and diversification. This review covers these approaches at a general, and hopefully accessible, level.

CONTINUOUS-TIME MARKOV CHAINS WITH A FINITE STATE SPACE

Many methods are based on this one general process, ranging from estimating DNA substitution rates on a tree to looking at the effect of foraging time on coat color. Continuous-time means that time can be broken up into arbitrarily small chunks (rather than being in discrete units like

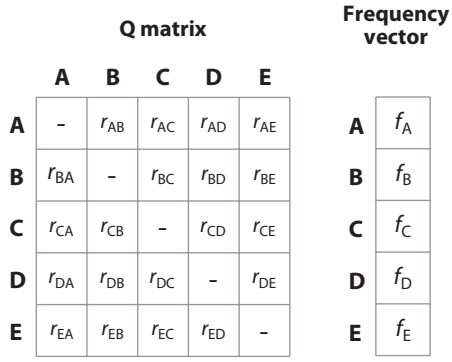

Figure 1

Basic structure of a continuous-time Markov chain finite state space (CTMC-FSS) model. The two main elements are an instantaneous rate matrix, **Q**, and a vector of state frequencies.

generations). A Markov chain is a random process in which the next step depends only on the present step, not steps further in the past. The probability of a nucleotide mutating from T to G when hit by a gamma ray does not depend on how long the position has been in state T or what state it was in before T. One common confusion is that while Markov chain Monte Carlo often appears in the context of Bayesian approaches in phylogenetics, there is nothing necessarily Bayesian about a Markov process. A finite state space means there are only certain states possible, such as A, T, G, or C for a particular site in DNA.

Figure 1 shows the basic elements of the model. When at a particular state, A, there is a probability that in a tiny time interval A becomes B. Call this rate r_{AB}. Such rates occur between every pair of states. They are often put into a matrix with the "from" state on the rows and the "to" state on the columns. This matrix is known as the instantaneous transition rate matrix, often given the symbol **Q**. The diagonal element r_{AA} is just the negative of the sum of the other elements, $-(r_{AB} + r_{AC} + r_{AD} + r_{AE})$. For convenience, this is usually just indicated as "−". There is also a vector of state frequencies. These can be optimized given a model set to fixed values (such as equal), set to empirical (observed) distributions, or set to equilibrium values. There is extensive work on the properties of continuous-time Markov chains, but to understand the links between methods, there are two things to remember. One is that while **Q** only has instantaneous rates, it is straightforward to get a matrix of probabilities of change between each state over a time interval t. This matrix is known as the transition-probability matrix and is frequently symbolized as $P(t)$ (P for probability, t because different length intervals have different transition probabilities). It is equal to $e^{\mathbf{Q}t}$. Take, for example, the Jukes-Cantor model of DNA evolution, where all transition rates are equal (all off-diagonal entries of **Q** are the same). On a short branch, the probability of starting at T and ending at A would be much smaller than the probability of starting at T and ending at T (not much time for change). But on a very long branch, the two probabilities would approach equality. Second, the vector of state probabilities at the end of a branch is the vector of probabilities at the beginning of a branch multiplied by $P(t)$. Pioneering work (Felsenstein 1981, Janson 1992, Pagel 1994, Yang 1994a) brought this model to tree inference and later to post-tree analysis of trait evolution.

The beauty of this model is that given a tree, transition rates, and frequencies at the root, one can calculate the probabilities of observed tip states. One could also examine different transition rates, root state frequencies, and even states elsewhere internally on the tree and see how they affect the probabilities of the tip states (the likelihood). A tree search involves optimizing some or all of

a GTR nucleotides

	A	G	C	T
A	-	r_{AG}	r_{AC}	r_{AT}
G	r_{AG}	-	r_{GC}	r_{GT}
C	r_{AC}	r_{GC}	-	r_{CT}
T	r_{AT}	r_{GT}	r_{CT}	-

b Binary correlation

	00	01	11	10
00	-	r_A	0	r_B
01	r_C	-	r_D	0
11	0	r_E	-	r_F
10	r_G	0	r_H	-

c Covarion

	A+	T+	A–	T–
A+	-	r_{AT}	δ	0
T+	r_{TA}	-	0	δ
A–	$k\delta$	0	-	0
T–	0	$k\delta$	0	-

d Ordered transitions

	0	1	2	3
0	-	r_{01}	0	0
1	r_{10}	-	r_{12}	0
2	0	r_{21}	-	r_{23}
3	0	0	r_{32}	-

Figure 2

Similarities between methods based on continuous-time Markov chain with a finite state space (CTMC-FSS) models. (*a*) General time reversible (GTR) nucleotide model; (*b*) model of correlated binary characters (Pagel 1994); (*c*) covarion model using variables from Penny et al. (2001), showing only two of four nucleotides for compactness; and (*d*) a model of ordered transitions (for example, state 1 can only go to its neighboring states, 0 or 2). Note that despite these models being intended for different kinds of characters (nucleotides, binary characters, and multistate characters) and for different purposes (tree inference, character correlation, examination of the process of hidden rate changes, and looking at transitions of a phenotypic character), they are restrictions of the same general **Q** matrix, with different rates set to equal each other or set to zero, and with the state names themselves varying. Not shown are the frequency vectors, which are also quite similar. If the frequencies are based on equilibrium values, they are specified by the **Q** matrix alone.

the above parameters as well as the tree itself: The basic model is the same (note that this review largely ignores methods intended just to infer the tree alone). Many Bayesian and likelihood-based models for discrete characters are just this general model with particular restrictions on the **Q** matrix. For example (**Figure 2**), when one infers a tree using a general time reversible (GTR) model, it is just a model where the **Q** matrix has four rows and columns (one for each nucleotide) and where rates are constrained such that the rate from state *i* to state *j* is forced to be equal to the rate from state *j* to state *i*, for any *i* and *j*. Pagel's model of binary character correlation (Pagel 1994) also has a four-by-four **Q** matrix: Each possible pair of binary states just becomes a single state in a four-state character, and the **Q** matrix is restricted so that changes in two binary states cannot happen simultaneously (the instantaneous rate of going from 00 to 11 is set to zero, so this transition must be accomplished by going through 01 or 10 intermediates). Seen this way, extensions for correlations of multiple, multistate characters is a straightforward change of mapping each possible combination to a different discrete state and making sure that instantaneous changes of multiple characters at once are prevented (making sure that there is enough data to fit the rates involved is a later empirical problem). A covarion model (Galtier 2001, Penny et al. 2001, Tuffley & Steel 1998), a model with characters invisibly converting between changeable and invariant modes, is also a restriction of the **Q** matrix (and if it had just two observed states, say A and T, as in **Figure 2**, it is also a four-by-four matrix; with four observed states, it would be an eight-by-eight matrix). Models that allow for different transition rates between different states for a single character, perhaps even treating characters as ordered (the character states form a chain), are restricted **Q** matrices, as are models of protein evolution that use a fixed rate matrix, such as a JTT (Jones et al. 1992) matrix. Statistical models for biogeography that allow species to occur in multiple discrete localities (Ree & Smith 2008) are also based on this basic model. The instantaneous rate parameters may be constrained in various ways (e.g., set to zero or set to equal one another) and may have different names (e.g, the rate of going from diurnal to nocturnal, the rate of changing from valine to lysine, the rate of entering into a variable covarion site, etc.), but the basic model still applies. Even maximum parsimony for discrete characters is an extension of the CTMC-FSS method that allows different rates on different branches for different characters

[the no common mechanism model (Tuffley & Steel 1997)], though the basis for its appeal is not that it fits a complex model but rather that it assumes change is rare. Some of the methods involving parsimony, such as the concentrated changes test (Maddison 1990), may be outside the CTMC-FSS framework, but this has yet to be formally evaluated. Sometimes even models that appear to be unbounded birth-death models (see below) are actually CTMC-FSS models. For example, Hahn et al. (2005) look at the evolution of gene family size up a tree. They treat gene family size as an ordered, discrete character that can only attain a particular maximum size (100 genes, in their initial work). There are then transitions on the tree from one number of genes to a neighboring number of genes, with the same gain and loss rate between neighboring states (same chance of going from 2 to 3 genes as from 82 to 83 genes), except with no gain rate from state 0 (lost forever) and state 100 (higher numbers of genes than 100 are prohibited).

Ancestral states can be estimated at the root of the tree or at internal nodes by using this same framework. This can be done for joint estimates (Yang et al. 1995) or marginal estimates. Joint estimates find the single best reconstruction across all nodes that maximizes the likelihood; this is like inferring someone's favorite meal (a joint collection of several food items). Marginal estimates find the single best reconstruction at each node examined alone, integrating over the possibilities at all other nodes; this is like inferring someone's favorite food item, averaging across all the meals in which it could be eaten. If a question involves a complete set of reconstructions, the joint approach is better; if the question involves a reconstruction at one particular node, a marginal approach is preferred. Stochastic character mapping (Huelsenbeck et al. 2003, Nielsen 2002) can use the **Q** matrix to evolve data up a tree, subject to the constraint that the data at the tips must match the observed data. Each sampled history is a single instance of a character evolving up a tree; one can summarize across these histories the number of times a character changes on a branch or which parts of a branch are in a particular character state. This is the only currently available method to reconstruct changes at points within branches, though evolutionary pathway likelihood (Steel & Penny 2000) as a reconstruction method of traits along branches of a fixed tree remains relatively unexplored.

Other statistical ideas can be added to this basic model. For example, using priors and estimating posterior probabilities converts it into a Bayesian approach (note that if other parameters are being estimated, such as tree topology, those must also be embedded in a Bayesian model). There are several examples of this (Pagel et al. 2004, Schultz & Churchill 1999). If there is a biased sample of characters such that invariant characters are not sampled, the bias can be corrected for in the model (Felsenstein 1992, Lewis 2001). Uncertainty in observations can also be directly dealt with in this model (Felsenstein 2004).

Heterogeneity can be incorporated through additions to this model (**Figure 3**). Heterogeneity can mean different processes at different characters (such as the difference in rates and state frequencies between nuclear and mitochondrial genes), different processes at different parts of the tree (early versus late in evolution, or woody versus herbaceous plants), or a combination of these. One approach to dealing with different processes at different characters is to use a partitioned model (Yang 1996). One set of characters gets assigned to one model, and a different set gets assigned to a different model. If the only difference is rates at different sites, a common method is to use a discrete gamma distribution (Yang 1994b). This divides a gamma distribution into k rates, and the overall likelihood for each site is the weighted sum of the site likelihoods with the same **Q** matrix but scaled by one of the k overall rates. The phylogenetic mixture model (Pagel & Meade 2004) works in a similar way, but rather than calculating the likelihood at a site using a single **Q** matrix linearly transformed by several rates to form a set of **Q** matrices, it calculates the likelihood at a site using multiple **Q** matrices (and also, potentially, different overall rates). For a given **Q** matrix it is the same as the basic CTMC-FSS; the innovation is using the sum

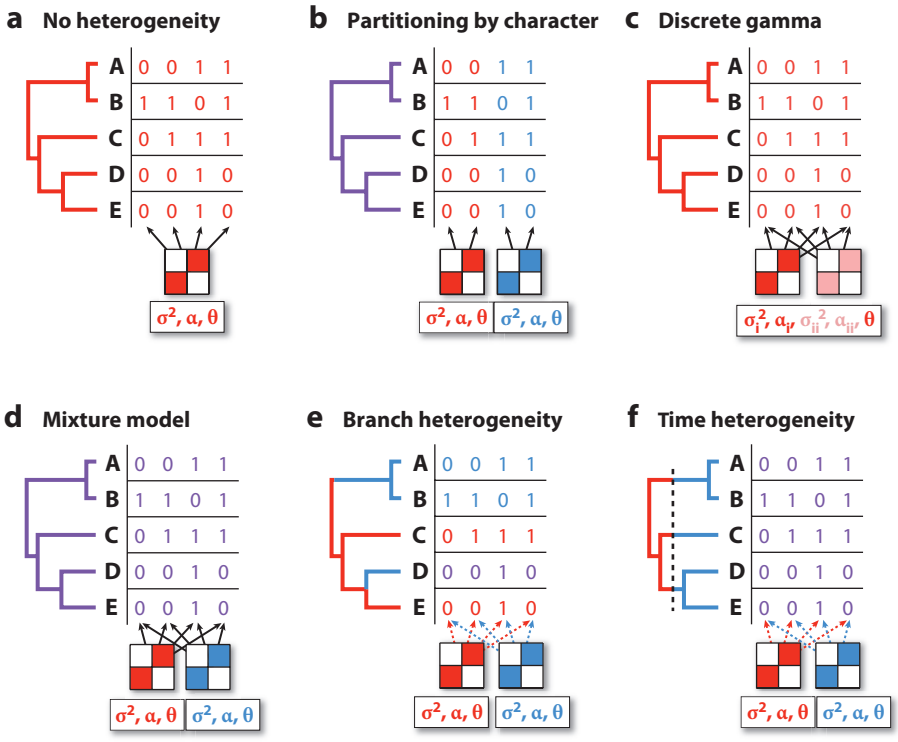

Figure 3

Dealing with heterogeneity on a tree. In each case, different models are represented in different colors (*red* and *blue*). Elements used with multiple models are in purple. Icons for continuous-time Markov chains with a finite state space (CTMC-FSS) models (**Q** matrix) and multivariate normal models (Ornstein-Uhlenbeck parameters σ^2, α, and θ) are shown for each to suggest how the same approach to heterogeneity can be used for each type of model. (*a*) The typical homogenous model. (*b*) Partitioning by character. Characters 1 and 2 (*red*) use one model, whereas characters 3 and 4 (*blue*) use another, but all share the same tree. (*c*) A discrete gamma model (Yang 1994b) with only two rate categories. With this approach the likelihood at each site is a weighted sum of likelihoods under different **Q** matrices, which are all linear transformations of the same basic matrix (thus showing the two models as very similar in *red* and *pink*). This has not yet been applied to continuous characters but could work in the same way. (*d*) A phylogenetic mixture model (Pagel & Meade 2004). This is like the discrete gamma model but allows for **Q** matrices that are not linear transformations of each other. (*e*) A branch heterogeneity model. Different branches may be assigned different models. This, and the next panel, can be done with birth-death models as well. (*f*) A time heterogeneity model. Different time periods are assigned to different models. Note that these various approaches to dealing with heterogeneity can be mixed with each other and with tree-stretching approaches.

of the likelihoods for the data under different **Q** matrices as the likelihood rather than using the likelihood under one **Q** matrix to account for differences between sites. Other approaches for dealing with rate heterogeneity across sites allow for rates to be correlated between neighboring sites (Felsenstein & Churchill 1996, Yang 1995). Heterogeneity across a tree has received less attention. Some have investigated allowing different state frequencies at different parts of the tree (Galtier & Gouy 1998, Yang & Roberts 1995). There has been some work on allowing different **Q** matrices on different, prespecified branches in order to evaluate different amounts of selection on different branches (Yang & Nielsen 2002, Zhang et al. 2005). It is important to realize that though many methods for dealing with heterogeneity envision using them for nucleotide data,

there is no necessary requirement that these are the only sorts of data that can be used because underneath them is the same CTMC-FSS model. For heterogeneity across character models, anything other than full partitioning may not make sense unless there is some biological reason to think that different characters would have similar transition matrices (which may occur if they are multiple characters of the same type such as gene presence/absence or secondary compound presence/absence). For heterogeneity across the tree, which can be used for even a single character, characters of many types could be examined. For example, one analysis could investigate whether the ratios of transition rates between herbivory, omnivory, and carnivory in mammals changed after the KT extinction by fitting different **Q** matrices to branches (and parts of branches) before and after the KT event.

MULTIVARIATE NORMAL DISTRIBUTION

Normal distributions are common in biology. By the central limit theorem of statistics, the sum of a set of independent random variables (each with finite mean and variance, but no stronger requirements) is normally distributed. The log of a product of positive independent random variables is also normally distributed. For example, if body mass in a species continues to be multiplied by various factors (cooler climate leads to a 10% increase in size, lower food availability leads to a 5% decrease in size, competition with other species leads to a 7% increase in size) across many replicates of this sort of evolutionary process, the log(body mass) should be normally distributed. This process is often called Brownian motion. Many evolutionary processes, ranging from tracking a selective optimum to genetic drift, can lead to this pattern (Edwards & Cavalli-Sforza 1964, Felsenstein 1988, Hansen & Martins 1996).

If looking at a single trait for a single species, that trait would be normally distributed under this model; that is, rerunning evolution many times would lead to a normal distribution of trait values. This distribution could be characterized by its mean and variance. If considering a pair of species, or a pair of traits for a single taxon, it becomes a bivariate normal distribution: Each element has its own mean and variance but there may also be covariance between the two elements. For multiple species and/or multiple traits, it becomes a multivariate normal distribution. Each element has its own mean and variance, but may also have covariance with the other elements (**Figure 4**). For simple Brownian motion of a single character, the variances are the root-to-tip length for each taxon (in units of time) multiplied by the rate of Brownian motion. The covariances are the shared time from the root of the tree to the most recent common ancestor of the pair of taxa multiplied by the Brownian motion rate. The character means are equal to the state at the root.

Some approaches allow particular branches of a tree to be assigned a priori to different rate categories for continuous characters (though the same approaches would work for discrete characters) and then allow the rates in each category to be estimated (O'Meara et al. 2006, Thomas et al. 2006), which is equivalent to stretching particular branches [in the noncensored approach of O'Meara et al. (2006)]. Recent advancements allow automated detection of the optimal locations for these rate parameters (Eastman et al. 2011, Thomas & Freckleton 2012). Ornstein-Uhlenbeck models are similar in that they stretch branch lengths but may also transform expectations. Brownian motion allows traits to wiggle, with the same expectation of variance in each unit of time. Ornstein-Uhlenbeck models are sometimes described as rubber band models: A species trait value wiggles through time but is drawn back to a mean value with some force. Hansen (1997) developed a model that allowed the mean value to change over the tree while keeping the wiggle and attraction parameters constant. Butler & King (2004) developed a likelihood implementation of this model. Beaulieu et al. (2012) developed a more general model that allowed mean, attraction, and wiggle to all vary either together or at different parts of the tree. A different use of this model

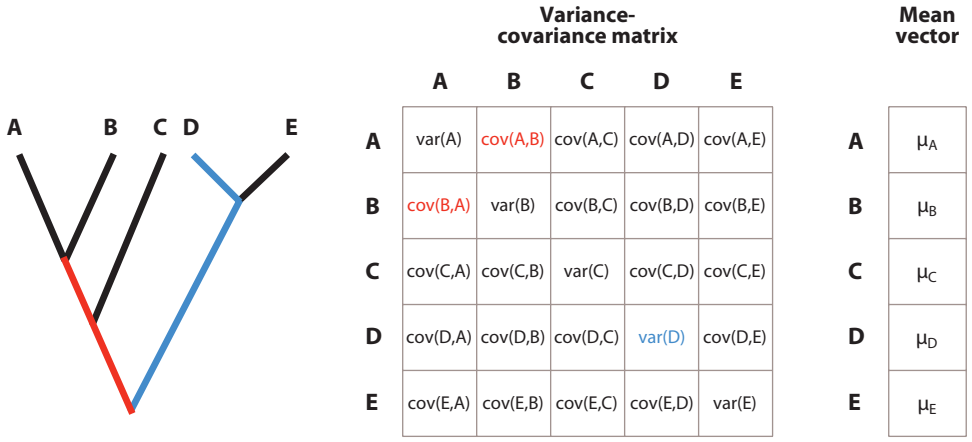

Figure 4

Multivariate normal distribution. The figure shows a tree, the tree's variance-covariance matrix, and the vector of means (which, under Brownian motion, would equal the root state). Highlighted are the branches leading to covariance between taxa A and B (*red*) and the branches leading to variance in D (*blue*).

is to look at correlations between characters. Revell & Collar (2009) developed an approach that estimates variances of and covariances between traits. Huelsenbeck & Rannala (2003) developed a model that can do that while also integrating uncertainty in the trees. Hansen et al. (2008) extend correlation approaches to allow Ornstein-Uhlenbeck processes. Ancestral states may be estimated under the multivariate normal process (Schluter et al. 1997) as they are with CTMC-FSS; in both cases, uncertainty in reconstructions can be substantial (Cunningham et al. 1998). The squared-change parsimony approach for continuous characters (Maddison 1991) recovers the same ancestral state estimates as Brownian motion models using a multivariate normal.

An interesting use of the multivariate normal process for discrete (and sometimes continuous) characters comes from Felsenstein's adoption of Wright's threshold model (Felsenstein 2005, 2012). A continuous trait, called the liability, evolves on the tree under a multivariate normal process. If this liability is above a certain threshold, a discrete character has one state; otherwise it has a different state. This can be used for multiple discrete and continuous characters (Felsenstein 2012). For the discrete character, the threshold model generates a different evolutionary pattern than would be expected under CTMC-FSS (for example, change in the discrete character is most likely immediately after a previous state change, where the liability is near the threshold, rather than constant through time), but the underlying process is multivariate normal.

BIRTH-DEATH BRANCHING PROCESSES

Finally, there are birth-death branching processes (Kendall 1948, Nee 2006). Each individual in a population has a certain probability of a birth (leading to two offspring) or a death (the individual ends). In the phylogenetic context, the individual is usually taken to be a species. In a pure birth model, also known as a Yule model (Yule 1924), the death rate is set to zero and the birth rate is constant through time. In other models, both rates are constant but allowed to be nonzero. It is also possible to vary the rates. Logistic growth, for example, is a pure birth model where the birth rate is a function of the number of individuals. There are many similarities between a birth-death branching process and a CTMC-FSS model, as they are both continuous-time Markov chains with discrete states. The main difference is that in a birth-death branching process there

is no maximum number of states. State zero is also an absorbing state in birth-death models (no speciation follows after all the species are extinct), which introduces an ascertainment bias, at least when working with neontological data: One only studies groups where some members have survived. It is the same problem as studying whether dolphins actually rescue drowning sailors. The reputed pattern may be due entirely to ascertainment bias: Dolphins might push drowning sailors in random directions, but only those sailors pushed shoreward survive to tell of the dolphins' influence. Such ascertainment bias needs to be considered by methods dealing with only surviving species. CTMC-FSS methods may also have absorbing states, but a more likely ascertainment bias in those methods as used in phylogenetics is looking at only variable characters (see discussion above).

One major focus of evolutionary investigations using trees is examining what diversification (speciation minus extinction) rates are on a tree. Unlike trait evolution, a main focus here has been investigating heterogeneity across a tree. Point estimates of a net diversification rate can be calculated given just the age of a clade and the number of taxa [reviewed by Magallon & Sanderson (2001)]. There are likelihood estimators for birth and death rates given a tree with branch lengths in units of time (Bokma 2008a, Nee et al. 1994), though death rates may be quite uncertain, especially if estimated based on just extant taxa (Rabosky 2010).

One question often addressed is how the diversification rate has changed over time. A common quantitative test uses the gamma test statistic (Pybus & Harvey 2000), though its performance has been criticized (Cusimano & Renner 2010, Liow et al. 2010). This is intended to measure speedups or slowdowns in diversification rates relative to a constant birth-death null. There are also approaches to fitting other models to trees, such as a density-dependent model of the number of species in a group (Rabosky 2006), a slowdown in speciation rate immediately after a speciation event (Losos & Adler 1995), or the use of survival models to allow flexible estimation of diversification rates (Paradis 1997).

Rather than comparing early versus later parts in the tree, other questions relate to comparing different clades in the tree. One simple such test is the Slowinski-Guyer test (Slowinski & Guyer 1989), which can determine whether two sister clades are of unusually unequal diversity. There have been several later modifications or improvements to this test (McConway & Sims 2004, Paradis 2012). Sanderson & Donoghue (1994) developed an early approach that allowed for different diversification rates on different branches. The SymmeTREE approach was designed to identify heterogeneity in rates across a whole tree (Chan & Moore 2002) and in particular parts of a tree (Moore et al. 2004). The MEDUSA approach (Alfaro et al. 2009) can find areas on a tree where clades are diversifying unusually slowly or quickly.

TREE STRETCHING

Using any of the three general frameworks above, especially the first two, one common approach is to look at heterogeneity in rates of evolution by doing tree stretching, which allows one to algorithmically transform a set of branch lengths on the tree. Many of the most common stretching models come from work by Pagel (Pagel 1997, 1999). A lambda transform affects the ratio of internal to terminal branch lengths. With $\lambda = 0$, change only happens on terminal branches. This can measure how much of evolutionary change is due to evolution happening at a constant rate through the tree and how much is due to only very recent changes. For example, if species were evolving a continuous trait that rapidly moved between fixed bounds, the influence of phylogenetic history would be erased and lambda would thus be close to zero. A kappa transform raises all branch lengths by the power κ. A value of 0 creates equal lengths for all branches, which matches a model where changes only happen after speciation events and where there are no missing speciation

events on the tree (a completely sampled tree with no extinctions). A delta transform stretches the tree to make changes more likely early (if $\delta < 1$) or later (if $\delta > 1$). Other transforms that may be applied include modifying branch lengths so that changes accelerate or decelerate through time [the ACDC model of Blomberg et al. (2003)] or modifying branch lengths so that branches after some threshold are a different length than branches before some threshold (Harmon et al. 2008). Some of the approaches in the Multivariate Normal Distribution section above (O'Meara et al. 2006, Thomas et al. 2006) also amount to simple stretching of prespecified parts of a tree. Note that any method that uses information on tree branch lengths can have tree stretching added to it, even if the particular stretching technique was developed for a different kind of data.

HYBRID METHODS

Historically, if a question were solely about trait evolution, differential diversification could be ignored. Many methods focused on the effect of traits on diversification itself would first reconstruct the tree, then reconstruct the traits, and only then examine the effect on diversification. However, in a key paper, Maddison (2006) showed that the distribution of a binary trait on a tree is affected by the effect of the trait's states both on diversification rates and on the transition rate between those states [note that this concern was raised fourteen years earlier by Janson (1992) but was largely ignored]. The traditional approach of separately looking at the transition and diversification aspects of the trait could provide the wrong answer. For example, imagine a trait where 0 to 1 transitions happen twenty times faster than 1 to 0 transitions. The number of taxa in state 1 should equilibrate at twenty times greater than the number of taxa in state 0. This could lead several methods for looking at diversification rates to estimate that the diversification rate in state 1 is greater than that in state 0 (thus leading to the increased number of taxa with state 1). In contrast, if state 1 does lead to higher diversification rates, this might be misinterpreted by a method looking at transition rates as a higher 0 to 1 than 1 to 0 rate. Maddison's (2006) work represents a change on the order of the change created by Felsenstein's independent contrasts paper (Felsenstein 1985), which showed that just conducting raw species contrasts might lead to misleading results because phylogeny is not taken into account. Community standards now generally expect a researcher looking into correlations between traits across many species to consider whether there is a phylogenetic effect that needs to be taken into account. We now know there could be an effect of differential diversification on estimates of trait evolution or an effect of unequal trait evolution on estimates of trait-associated diversification. Investigation of these effects is of growing importance. Note that this process need not only occur with key innovations or rapidly radiating groups: It could happen any time different states of a trait may be correlated (even via trait correlation with a second, unmeasured trait) with different extinction and/or speciation rates. For example, if carnivores tend to have shorter intestinal tracts than herbivores, and carnivory tends to lead to higher extinction rates (as carnivores are higher up the trophic pyramid), then a reconstruction of intestinal length ignoring potential correlation with speciation and extinction rates may be misleading. This happens even though there is no credible causal link between intestines and speciation or extinction: Lack of causation does not imply lack of correlation.

Fortunately, there are methods to deal with this. The first such method, BiSSE (Maddison et al. 2007), uses a model for binary traits that allows different states to have different transition rates, birth rates, and death rates. This requires fully resolved and sampled phylogenies. A later extension relaxed this requirement (FitzJohn et al. 2009), and other work by FitzJohn has extended it to multistate characters (FitzJohn 2012). There is now an approach for continuous characters (QuaSSE) that can deal with a single trait that evolves under a diffusion process (like Brownian motion or the Ornstein-Uhlenbeck process) but that also may affect diversification rate (FitzJohn 2010).

There is a model, GeoSSE, for dealing with geography, where a species may occur in multiple localities and where diversification rate may be tied to locality (Goldberg et al. 2011). There are not, yet, methods that deal with multiple characters (though the multistate character approach can deal with this if characters are recoded) or with heterogeneity over the tree, though such methods will undoubtedly be developed. Bokma (2002, 2008b) independently developed methods that examine the joint effect of diversification, missing species, and change at nodes and along branches.

OTHER KEY METHODS

Although many methods fit in the above three broad categories, many methods, including some of the most important ones, do not. For example, there is a rich array of regression-based methods that correct for phylogeny as an error term or as a correlation term. These include methods such as generalized linear models (Martins & Hansen 1997), generalized estimating equations (Paradis & Claude 2002), the phylogenetic mixed method (Housworth et al. 2004, Lynch 1991), phylogenetic generalized least squares, and phylogenetic autocorrelation (Cheverud et al. 1985, Gittleman & Kot 1990) among others (e.g., Ives & Garland 2010). They are too flexible to put into one category, but they are often essential tools, especially when examining data of mixed types. Independent contrasts (Felsenstein 1985), a method to correct for nonindependence of data on a tree, is consistent with the multivariate normal framework but works slightly differently (using estimates only incorporating information from descendant taxa, for example). There is a growing array of methods to use for phylogeographic inferences (history of population subdivision, history of population size, and so forth) that use richer models than any of the above (Knowles 2009). Some biogeographic models (Ronquist & Sanmartín 2011) and other coevolutionary models (Page 1994) also use methods beyond what was discussed here. There is also a wide array of randomization techniques for things like community phylogenetics (Webb et al. 2002) or phylogenetic signal (Blomberg et al. 2003). Methods that simulate based on actual mechanisms are growing in importance, as well. For example, Pigot et al. (2010) developed a model that simulates range changes and vicariance events explicitly, which generates distributions different from those seen under birth-death models. The growth of approximate Bayesian computation (ABC) methods (Beaumont et al. 2002, Huang et al. 2011, Rabosky 2009) is making the use of explicit models much more feasible.

PARAMETER ESTIMATION, HYPOTHESIS REJECTION, AND MODEL SELECTION

Practitioners of evolutionary inference using phylogenies differ in their statistical approaches. There are of course the differences between those who use explicit models and those who do not as well as between Bayesians and frequentists. However, given that many methods are not implemented in software (see the sidebar, Software) adopting all these perspectives, many empiricists care more about using a method that provides answers to their questions than about fealty to a particular statistical worldview. However, there is a more basic argument that affects which methods are developed and how they are applied. It is the ongoing debate about using methods to answer biological questions by estimating parameters versus using methods to reject null hypotheses (Stephens et al. 2007). A simple example of the latter is the Slowinski & Guyer (1989) test. It returns whether the distribution of diversity in two sister clades is more uneven than expected under a basic null model. There is utility to this: Learning that a clade of 1,000 taxa is not significantly more diverse than its sister clade having 30 taxa helps correct our perceptions of

SOFTWARE

Methods for evolutionary inferences using phylogenetics need to be implemented in software in order to be used. There are two notable recent trends in software for evolutionary inferences. The first is the move to open source: Rather than being a black box of compiled computer code, the human-readable code underlying the program is distributed freely for inspection and modification. Most new evolutionary methods are now released in open-source software.

The other major trend is the growth of R (R Dev. Core Team 2011) as a framework for methods for evolutionary inference. One of its main advantages is its use of packages to add functionality. For example, the APE package (Paradis et al. 2004) can load, traverse, and display trees, so someone developing a new method can use those existing functions. There are over thirty packages devoted to phylogenetics alone (O'Meara 2012), as well as a book (Paradis 2011). R is certainly not the only option. For example, Mesquite (Maddison & Maddison 2007) is a Java application that implements an abundance of methods and that can be extended with modules. There are also projects to build bioinformatics libraries in languages such as Python, Perl, and C++.

what is expected under null models. Likelihood ratio tests can be used with approaches like tree stretching to show that simple Brownian motion is rejected. Rejection of a null model is based on two factors, how well the data fit the null model and data set size. As data sets grow larger, it will become increasingly easy to reject basic null models: the chances of a clade of 10,000 taxa evolving traits exactly under simple Brownian motion, or having a completely constant diversification rate through time, are low. Rather than just inferring whether a parameter is statistically significantly different from zero, one must determine whether it is statistically and biologically significantly different from zero.

One response to criticisms of null hypothesis testing has been to adopt information theoretic approaches, such as the Akaike Information Criterion (Akaike 1973, Burnham & Anderson 2002). These approaches attempt to identify the model that minimizes information lost about nature. A model that is too simple may have missed important factors that should be included, whereas a model that is too complex may have too many parameters to fit well. This approach, like approaches that involve Bayesian model evaluations, lends itself well to multimodel inference: drawing conclusions based on a set of weighted models, rather than putting all faith in a single best model. The risk, however, is in interpreting model selection as hypothesis testing: treating a simple model with low weight as a rejected model and making that a focus.

Another recent advance has been the reintroduction of the idea of model adequacy. Existing approaches to model fitting [likelihood ratio tests, information theoretic measures, Bayes factors, reversible jump MCMC (Huelsenbeck et al. 2004, Pagel & Meade 2006), and so forth] estimate which model is least bad. There is little information from such approaches about whether a model is actually a good fit to the data [with the noteworthy exception of methods that are based on regression ideas (Martins & Hansen 1997) or various approaches to make sure contrasts are well-standardized (Garland et al. 1992)]. An approach to looking at whether the model adequately reflects the data is to simulate data under the inferred model and see if the simulated data are indistinguishable from the observed data by some measures. This has been done recently for comparative methods by Boettiger et al. (2011) but was done earlier for substitution models (Goldman 1993) and for looking at conflicting phylogenetic signal (Huelsenbeck & Bull 1996). In a Bayesian context, a posterior predictive distribution can be used in much the same way (Huelsenbeck et al. 2001). Doing this sort of test of adequacy will be important in developing more

realistic models: comparing the fits of models will only tell which is better, but by showing that in some situations all are inadequate, researchers will know where to begin work on better models.

TREE SIZE AND HETEROGENEITY

In less than twenty years, large tree data sets have gone from 500 taxa (Chase et al. 1993) to over 70,000 taxa (Goloboff et al. 2009). The number of free parameters that can be accurately estimated has also increased, though the exact bounds of this remain to be explored. For example, if a reasonable rule of thumb of 100 observations per parameter is used, this suggests that a model with 700 free parameters could be estimated from a 70,000-taxon data set. In contrast, the commonly used Brownian motion model has just two parameters, the rate and the root state. It is an empirical question, but it seems likely that complex models with perhaps as many as dozens of free parameters may be appropriate at this scale. More importantly, as greater sections of the tree of life are used in analyses, the assumptions of the homogeneity of processes used in many models begin to break down. A phylogenetically corrected regression of brain size on body size may work well in New World monkeys, but as one expands the set of taxa to include all primates, then all mammals, then all vertebrates, the reasonableness of the model no longer holds. The factors linking brain and body size in arboreal mammals may be very different from those acting in teleost fishes or birds, as traits like aquatic environment, flight constraints, social structure, diet, and other factors may vary over the tree. Doing this analysis phylogenetically remains important, but a more complex model than a single fixed regression may be justified. More than just a complicating factor, this promises new discoveries about biology as the power to perform analyses with more realistic models grows. There have been initial approaches to try different models on different parts of trees using a priori assignment (Beaulieu et al. 2012, Butler & King 2004, O'Meara et al. 2006, Thomas et al. 2006) as well as automatic placement of models (Alfaro et al. 2009, Eastman et al. 2011, Thomas & Freckleton 2012), but this will represent an important area of growth in the future.

CONCLUSION

As species evolve through time, information about their traits and changes is passed on to their descendants. Although much of this information is erased by subsequent changes, much still persists: The ancestor of bats, pterosaurs, and cows had four limbs; the squid, human, and fly lineages have maintained a homologous gene for initiating a light detector; and angiosperms have a higher diversification rate than many other plant groups. The combination of trait data, a phylogeny, and various methods can be used to make inferences about past and ongoing evolutionary processes. Although there are scores of different methods for making inferences about evolutionary processes using phylogenies, many can be reduced to a few essential models. This does two important things. First, it allows models to be more readily understood: Once a GTR DNA model is grasped, one also then understands a model for evolution of flight as a discrete trait. Second, it allows advances in approaches where one set of questions can be easily ported to a different set of questions. For example, the discrete gamma approach to rate heterogeneity (Yang 1994b) has been applied to nucleotide data for tree inference, but it could be applied to data on expression/nonexpression of various genes in developing flowers in a more realistic model for ancestral state reconstruction of flower development. The only real difference between the models is whether the states are called A, T, G, and C at a site or presence/absence of expression for a gene.

We are in a golden age of phylogenetics. The number of taxa and data set size for phylogenetic trees are expanding, yet there are still discoveries about relationships to be made. Computational

power on individual machines continues to grow as does the availability of free (to the community) supercomputing for phylogenetics (Goff et al. 2011, Miller et al. 2010). Data sets and trees are stored in accessible databases allowing reuse. This all points to more data allowing for more models, perhaps with more complexity and realism. As model adequacy is used more often, biologists will learn where existing methods are insufficient and will create new, more powerful approaches.

SUMMARY POINTS

1. There are a wide variety of methods for drawing evolutionary inferences given phylogenies, but despite this diversity, a few major models recur in many available methods.
2. Continuous-time Markov chain models with finite states underlie many diverse models, from DNA evolution to gene family size or biogeography.
3. The multivariate normal distribution underlies many methods for continuous data and at least one for discrete data.
4. Birth-death processes play major roles in inferences about speciation and extinction dynamics.
5. Tree stretching is a broad category of methods that allow tests of rate heterogeneity across time or taxa.
6. Methods that jointly estimate transition and diversification rates represent an important advance in comparative methods and may be of growing influence.

FUTURE ISSUES

1. Model adequacy will become more important as a way to measure how well models describe biological data.
2. As the sizes of trees grow, methods will increasingly have to deal with heterogeneity in processes at different parts of the tree.
3. Current models often describe patterns that may be caused by multiple processes. Mechanistic models may grow in importance, especially as simulation-based approaches become more accessible.
4. Most current methods assume species have a single trait value per character, with some important exceptions (Felsenstein 2005, 2008; Ives et al. 2007). Just as the realization of the importance of variation in gene histories is affecting tree reconstruction, appreciation of the variation in trait values within species will generate a need for further advances.
5. In many groups the true evolutionary history is reticulate, not fully tree-like. Methods for making evolutionary inferences on networks still need to be developed.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Support for some of the research leading to this review came from a fellowship to B.C.O. from the National Evolutionary Synthesis Center (NESCent), NSF #EF-0905606, and working group support from the iPlant Collaborative, NSF #DBI-0735191. I thank Barb Banbury, Jeremy Beaulieu, Martin Ryberg, Stacey Smith, and Peter Wainwright for feedback and guidance. Conversations with Joe Felsenstein contributed to an appreciation of the importance of dealing with heterogeneity of processes over a tree.

LITERATURE CITED

Akaike H. 1973. Information theory as an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, ed. BN Petrov, F Csaki, pp. 267–81. Budapest: Akad. Kiado

Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, et al. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Natl. Acad. Sci. USA* 106:13410–14

Barker D, Pagel M. 2005. Predicting functional gene links from phylogenetic-statistical analyses of whole genomes. *PLoS Comput. Biol.* 1:24–31

Beaulieu JM, Jhwueng D-C, Boettiger C, O'Meara BC. 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* 66:2369–83

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–35

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* 57:717–45

Boettiger C, Coop G, Ralph P. 2011. Is your phylogeny informative? Measuring the power of comparative methods. *Evolution* 66:2240–51

Bokma F. 2002. Detection of punctuated equilibrium from molecular phylogenies. *J. Evol. Biol.* 15:1048–56

Bokma F. 2008a. Bayesian estimation of speciation and extinction probabilities from (in)complete phylogenies. *Evolution* 62:2441–45

Bokma F. 2008b. Detection of "punctuated equilibrium" by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. *Evolution* 62:2718–26

Burnham KP, Anderson DR. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer

Butler MA, King AA. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164:683–95

Chan KMA, Moore BR. 2002. Whole-tree methods for detecting differential diversification rates. *Syst. Biol.* 51:855–65

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, et al. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Ann. Mo. Bot. Gard.* 80:528–80

Cheverud JM, Dow MM, Leutenegger W. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution* 39:1335–51

Cunningham CW, Omland KE, Oakley TH. 1998. Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol. Evol.* 13:361–66

Cusimano N, Renner SS. 2010. Slowdowns in diversification rates from real phylogenies may not be real. *Syst. Biol.* 59:458–64

Eastman JM, Alfaro ME, Joyce P, Hipp AL, Harmon LJ. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution* 65:3578–89

Edwards AWF, Cavalli-Sforza LL. 1964. Reconstruction of evolutionary trees. In *Phenetic and Phylogenetic Classification*, Vol. 6, ed. VH Heywood, J McNeill, pp. 67–76. London: Syst. Assoc. Publ.

Felsenstein J. 1981. Evolutionary trees from DNA sequences—a maximum-likelihood approach. *J. Mol. Evol.* 17:368–76

Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15

Felsenstein J. 1988. Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* 19:445–71
Felsenstein J. 1992. Phylogenies from restriction sites—a maximum-likelihood approach. *Evolution* 46:159–73
Felsenstein J. 2004. *Inferring Phylogenies.* Sunderland, MA: Sinauer
Felsenstein J. 2005. Using the quantitative genetic threshold model for inferences between and within species. *Philos. Trans. R. Soc. B* 360:1427–34
Felsenstein J. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *Am. Nat.* 171:713–25
Felsenstein J. 2012. A comparative method for both discrete and continuous characters using the threshold model. *Am. Nat.* 179:145–56
Felsenstein J, Churchill GA. 1996. A Hidden Markov Model approach to variation among sites in rate of evolution. *Mol. Biol. Evol.* 13:93–104
FitzJohn RG. 2010. Quantitative traits and diversification. *Syst. Biol.* 59:619–33
FitzJohn RG. 2012. *Diversitree: comparative phylogenetic tests of diversification. R package version 0.9–1.* **http://www.zoology.ubc.ca/prog/diversitree/**
FitzJohn RG, Maddison WP, Otto SP. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* 58:595–611
Galtier N. 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. *Mol. Biol. Evol.* 18:866–73
Galtier N, Gouy M. 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* 15:871–79
Garland T Jr, Harvey PH, Ives AR. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* 41:18–32
Gittleman JL, Kot M. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. *Syst. Biol.* 39:227–41
Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, et al. 2011. Frontiers: The iPlant Collaborative: cyberinfrastructure for plant biology. *Front. Plant Sci.* 2:34
Goldberg EE, Lancaster LT, Ree RH. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol.* 60:451–65
Goldman N. 1993. Simple diagnostic statistical tests of models for DNA substitution. *J. Mol. Evol.* 37:650–61
Goloboff PA, Catalano SA, Marcos Mirande J, Szumik CA, Salvador Arias J, et al. 2009. Phylogenetic analysis of 73 060 taxa corroborates major eukaryotic groups. *Cladistics* 25:211–30
Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* 15:1153–60
Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–51
Hansen TF, Martins EP. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–17
Hansen TF, Pienaar J, Orzack SH. 2008. A comparative method for studying adaptation to a randomly evolving environment. *Evolution* 62:1965–77
Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24:129–31
Housworth EA, Martins EP, Lynch M. 2004. The phylogenetic mixed model. *Am. Nat.* 163:84–96
Huang W, Takebayashi N, Qi Y, Hickerson MJ. 2011. MTML-msBayes: approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. *BMC Bioinformatics* 12:1
Huelsenbeck JP, Bull JJ. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. *Syst. Biol.* 45:92–98
Huelsenbeck JP, Larget B, Alfaro ME. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol. Biol. Evol.* 21:1123–33
Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic mapping of morphological characters. *Syst. Biol.* 52:131–58
Huelsenbeck JP, Rannala B. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution* 57:1237–47

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–14
Ives AR, Garland T Jr. 2010. Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.* 59:9–26
Ives AR, Midford PE, Garland T Jr. 2007. Within-species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.* 56:252–70
Janson CH. 1992. Measuring evolutionary constraints: a Markov model for phylogenetic transitions among seed dispersal syndromes. *Evolution* 46:136–58
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8:275–82
Kendall DG. 1948. On the generalized "birth-and-death" process. *Ann. Math. Stat.* 19:1–15
Knowles LL. 2009. Statistical phylogeography. *Annu. Rev. Ecol. Evol. Syst.* 40:593–612
Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50:913–25
Liow LH, Quental TB, Marshall CR. 2010. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? *Syst. Biol.* 59:646–59
Losos JB, Adler FR. 1995. Stumped by trees? A generalized null model for patterns of organismal diversity. *Am. Nat.* 145:329–42
Lynch M. 1991. Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45:1065–80
Maddison WP. 1990. A method for testing the correlated evolution of two binary characters: Are gains or losses concentrated on certain branches of a phylogenetic tree? *Evolution* 44:539–57
Maddison WP. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Syst. Zool.* 40:304–14
Maddison WP. 2006. Confounding asymmetries in evolutionary diversification and character change. *Evolution* 60:1743–46
Maddison WP, Maddison DR. 2007. *Mesquite: a modular system for evolutionary analysis. Version 2.0.* **http://mesquiteproject.org**
Maddison WP, Midford PE, Otto SP. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56:701–10
Magallon S, Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55:1762–80
Martins EP, Hansen TF. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* 149:646–67
McConway KJ, Sims HJ. 2004. A likelihood-based method for testing for nonstochastic variation of diversification rates in phylogenies. *Evolution* 58:12–23
Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Presented at *Gateway Comput. Environ. Workshop (GCE) Nov. 14, New Orleans, LA*, pp. 1–8
Moore BR, Chan KMA, Donoghue MJ. 2004. Detecting diversification rate variation in supertrees. In *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life*, ed. ORP Binida-Emonds, pp. 487–533. Dordrecht: Kluwer Acad.
Nee S. 2006. Birth-death models in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* 37:1–17
Nee S, Holmes EC, May RM, Harvey PH. 1994. Extinction rates can be estimated from molecular phylogenies. *Philos. Trans. R. Soc. Lond. Ser. B* 344:77–82
Nielsen R. 2002. Mapping mutations on phylogenies. *Syst. Biol.* 51:729–39
O'Meara BC. 2012. *CRAN task view: phylogenetics. Version 2012-02-02.* **http://cran.r-project.org/web/views/Phylogenetics.html**
O'Meara BC, Ane C, Sanderson MJ, Wainwright PC. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922–33
Page RDM. 1994. Parallel phylogenies: reconstructing the history of host-parasite assemblages. *Cladistics* 10:155–73
Pagel M. 1994. Detecting correlated evolution on phylogenies—a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. Ser. B* 255:37–45
Pagel M. 1997. Inferring evolutionary processes from phylogenies. *Zool. Scr.* 26:331–48

Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–84
Pagel M, Meade A. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* 53:571–81
Pagel M, Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* 167:808–25
Pagel M, Meade A, Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* 53:673–84
Paradis E. 1997. Assessing temporal variations in diversification rates from phylogenies: estimation and hypothesis testing. *Proc. R. Soc. Lond. Ser. B* 264:1141–47
Paradis E. 2011. *Analysis of Phylogenetics and Evolution with R*. Berlin: Springer-Verlag. 2nd ed.
Paradis E. 2012. Shift in diversification in sister-clade comparisons: a more powerful test. *Evolution* 66:288–95
Paradis E, Claude J. 2002. Analysis of comparative data using generalized estimating equations. *J. Theor. Biol.* 218:175–85
Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–90
Penny D, McComish BJ, Charleston MA, Hendy MD. 2001. Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *J. Mol. Evol.* 53:711–23
Pigot AL, Phillimore AB, Owens IPF, Orme CDL. 2010. The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. *Syst. Biol.* 59:660–73
Pybus OG, Harvey PH. 2000. Testing macroevolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. Lond. Ser. B* 267:2267–72
R Dev. Core Team. 2011. *R: A Language and Environment for Statistical Computing*, Vol. 1. Vienna, Austria: R Found. Stat. Comput.
Rabosky DL. 2006. Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* 60:1152–64
Rabosky DL. 2009. Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. *Syst. Biol.* 58:629–40
Rabosky DL. 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution* 64:1816–24
Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57:4–14
Revell LJ, Collar DC. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution* 63:1090–100
Ronquist F, Sanmartín I. 2011. Phylogenetic methods in biogeography. *Annu. Rev. Ecol. Evol. Syst.* 42: 441–64
Ryan MJ, Rand AS. 1995. Female responses to ancestral advertisement calls in túngara frogs. *Science* 269:390–92
Sanderson MJ, Donoghue MJ. 1994. Shifts in diversification rate with the origin of angiosperms. *Science* 264:1590–93
Schluter D, Price T, Mooers AO, Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51:1699–711
Schultz TR, Churchill GA. 1999. The role of subjectivity in reconstructing ancestral character states: a Bayesian approach to unknown rates, states, and transformation asymmetries. *Syst. Biol.* 48:651–64
Slowinski JB, Guyer C. 1989. Testing the stochasticity of patterns of organismal diversity: an improved null model. *Am. Nat.* 907–21
Steel M, Penny D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Mol. Biol. Evol.* 17:839–50
Stephens PA, Buskirk SW, del Rio CM. 2007. Inference in ecology and evolution. *Trends Ecol. Evol.* 22:192–97
Thomas GH, Freckleton RP. 2012. MOTMOT: models of trait macroevolution on trees. *Methods Ecol. Evol.* 3:145–51
Thomas GH, Freckleton RP, Székely T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proc. R. Soc. B* 273:1619–24
Tuffley C, Steel M. 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bull. Math. Biol.* 59:581–607
Tuffley C, Steel M. 1998. Modeling the covarion hypothesis of nucleotide substitution. *Math. Biosci.* 147:63–91

Vijaykrishna D, Smith GJD, Pybus OG, Zhu H, Bhatt S, et al. 2011. Long-term evolution and transmission dynamics of swine influenza A virus. *Nature* 473:519–22

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. 2002. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* 33:475–505

Yang ZH. 1994a. Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. *Syst. Biol.* 43:329–42

Yang ZH. 1994b. Maximum-likelihood phylogenetic estimation from DNA sequences with variable rates over sites—approximate methods. *J. Mol. Evol.* 39:306–14

Yang ZH. 1995. A space-time process model for the evolution of DNA sequences. *Genetics* 139:993–1005

Yang ZH. 1996. Maximum-likelihood models for combined analyses of multiple sequence data. *J. Mol. Evol.* 42:587–96

Yang ZH, Kumar S, Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141:1641–50

Yang ZH, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19:908–17

Yang ZH, Roberts D. 1995. On the use of nucleic acid sequences to infer early branchings in the tree of life. *Mol. Biol. Evol.* 12:451–58

Yule GV. 1924. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philos. Trans. R. Soc. Lond. B* 213:21–87

Zhang JZ, Nielsen R, Yang ZH. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22:2472–79

Annual Review of Ecology, Evolution, and Systematics

Volume 43, 2012

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Ecology, Evolution, and Systematics* articles may be found at http://ecolsys.annualreviews.org/errata.shtml