Review TRENDS in Ecology and Evolution Vol.19 No.9 September 2004

Phylogenetics Series

Bayesian inference of character evolution

Fredrik Ronquist

Computational Science and Information Technology, Florida State University, Tallahassee, FL 32306–4120, USA

Much recent progress in evolutionary biology is based on the inference of ancestral states and past transformations in important traits on phylogenetic trees. These exercises often assume that the tree is known without error and that ancestral states and character change can be mapped onto it exactly. In reality, there is often considerable uncertainty about both the tree and the character mapping. Recently introduced Bayesian statistical methods enable the study of character evolution while simultaneously accounting for both phylogenetic and mapping uncertainty, adding much needed credibility to the reconstruction of evolutionary history.

Evolution is a difficult phenomenon to study. It is rarely fast enough to be observed directly and only in exceptional cases is it possible to find physical evidence, such as fossils or ancient DNA, of past states and events. Fortunately, evolution leaves its footprint in the distribution of traits among living things. By studying this footprint, we can infer how organisms originated through the successive splitting of ancestral lineages, a process depicted in phylogenetic trees. Given a phylogenetic tree, we can also reconstruct the evolutionary history of individual traits of interest.

The wide range of questions that can be addressed by the INFERENCE (see Glossary) of ancestral states or paths of change in key traits on phylogenetic trees is fascinating. A few examples include the design of vaccines [1], the reconstruction of ancestral hormone receptors [2] and ancestral metabolic pathways [3], the inference of ancient behaviours [4], the identification of past dispersal patterns [5–7], the study of positive selection in proteins [8], the discovery of viral infection pathways [9], and the recognition of character correlation in coevolving lineages [10].

Many of these applications still rely on explicit or implicit PARSIMONY mapping of characters onto a single phylogenetic tree. The parsimony method finds the reconstruction that implies the smallest number of changes on the given tree; the solution is often intuitively obvious (Figure Ia in Box 1). The inferred ancestral states and character changes using parsimony typically reveal the process of evolution in exhilarating detail.

It has long been recognized that this approach ignores two important sources of error. First, the parsimony principle singles out the solution(s) requiring the minimum amount of change on the given tree, although there is usually a range of alternative reconstructions on the same tree that are almost as likely [11] (MAPPING UNCERTAINTY; Box 1). Second, the tree is almost never known without error [12] (PHYLOGENETIC UNCERTAINTY; Box 2). If there is a range of plausible trees, it is possible that the evolutionary history of a trait could differ depending on the tree. Clearly, ignoring either of these sources of error is potentially misleading.

Glossary

Bayesian inference: theory of statistical inference based on the idea of rational accumulation of scientific knowledge. Statistical models and model parameters are regarded as random variables, and statistical analysis uses data (observations) to update a prior probability distribution on these parameters to a posterior probability distribution.

Bootstrapping (nonparametric): procedure for examining the uncertainty in a statistical estimate by drawing new samples (pseudosamples) from the original sample, and repeating the statistical procedure for each of these new samples. There is also a parametric variant that generates new samples by using a parametric model estimated from the original sample.

Conditional probability: the probability conditioned on (given) some information; we can think of it as a relative probability. In Box 1, the conditional (relative) probability of ancestor B being purple (state 0) is $\Pr(B=0) = 0.00024/0.00037 = 0.65$. Hence, the conditional probability it being green is $\Pr(B=1) = 1 - \Pr(B=0) = 1 - 0.65 = 0.35$. Box 1

Inference: to draw conclusions about a statistical model using empirical data.

Likelihood: probability that a particular model (with specific parameter values) produced some observed data. For instance, the likelihood (probability) of the data in Box 1 is $L = 0.00037$ given the binary Markov model with $\pi_1 = 0.5$ and summing over ancestral states. If ancestor B has state 0 (purple), the likelihood is $L(B=0) = 0.00024$; if it has state 1 (green), the likelihood is $L(B=1) = L - L(B=0) = 0.00037 - 0.00024 = 0.00013$. Box 1

Mapping uncertainty: the error associated with reconstructing the evolution of a character on a given phylogenetic tree.

Markov chain Monte Carlo (MCMC): stochastic simulation technique for generating a sample from a complex distribution that is known up to a normalizing constant. It is widely used to sample Bayesian posterior distributions, where it is based on specially designed Markov models (similar but more complex than the ones used to model evolution; Box 3) and their tendency to move towards a stationary condition. Box 3

Maximum likelihood (ML): widely used method of statistical inference that finds the parameter values that maximize likelihood. For instance, when $\pi_1 = 0.5$, the ML state of ancestor B (Figure Ib in Box 1) is 0 (purple) because $L(B=0) > L(B=1)$. More typically, ML is used to estimate the free parameters of a probability model. For instance, if we vary π_1, we discover that the likelihood of the observed data is maximized when $\pi_1 \approx 0.20$. This is the ML estimate of π_1. Figure I, Box 1

Parsimony: inference principle based on minimizing cost; in evolutionary inference, usually the same as minimizing the number of character changes.

Phylogenetic uncertainty: the uncertainty in reconstructing character evolution owing to error in the phylogenetic estimate.

Posterior (probability distribution): probability distribution describing the knowledge about a model and its parameters after a Bayesian analysis. Can be used as the prior in a subsequent Bayesian analysis.

Prior (probability distribution): probability distribution specifying the knowledge about a model and its parameters before a Bayesian analysis.

Corresponding author: Fredrik Ronquist (ronquist@csit.fsu.edu).
Available online 21 July 2004

 0169-5347/$ - see front matter doi:10.1016/j.tree.2004.07.002

Box 1. Mapping uncertainty

We are interested in inferring the ancestral states of a character with two states, purple (0) or green (1). The states could, for example, represent particular behaviours, life-history traits or morphological features. The tree and the ages of the nodes are known; the scale is in amount of expected change.

Parsimony (Figure Ia) finds the reconstruction requiring the minimum amount of change between green and purple. In this case, there are two changes [marked (i) and (ii)] assuming gains and losses count equally. Ancestors are inferred as being either green or purple, but we do not know how certain these conclusions are. That is, we have not taken mapping uncertainty into account.

Likelihood analysis requires that we know the relative rates of $0 \rightarrow 1$ changes (π_1) and $1 \rightarrow 0$ changes (π_0). Assuming that these rates are equal ($\pi_0 = \pi_1 = 0.5$), for instance, we can calculate the probability of each ancestor being either green or purple (Figure Ib). The ancestral state is uncertain for ancestors A, B and F because they are on long branches or close to regions of the tree where a state change is likely.

In Bayesian inference, π_1 does not have to be fixed. Instead, we specify a prior probability distribution on π_1. In the absence of background information, we can assume that all possible values are equally likely (Figure Ic: prior). This enables us to infer ancestral states while weighting each π_1 value according to its probability given the data (Figure Ic: posterior). In our example, Bayesian inference simply adds a dash more uncertainty to the conditional probability values (Figure Id; the effect is most notable for ancestor C).

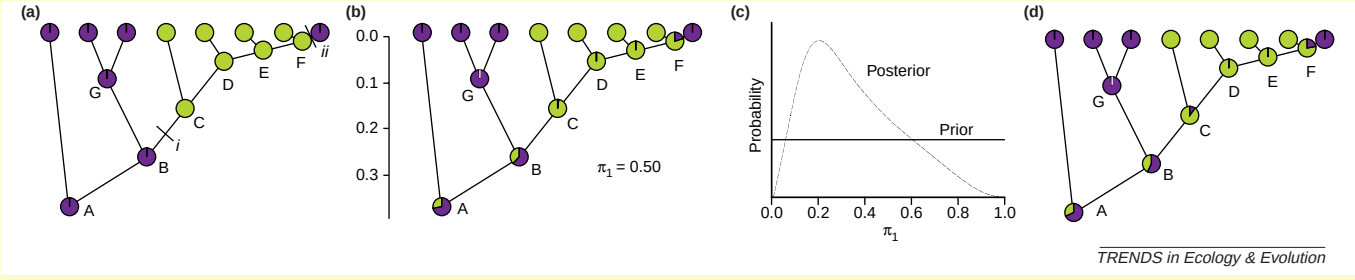

Figure I.

Sensitivity analysis

A simple way of examining the robustness of evolutionary inference is to look at how sensitive the results are to slight changes in the analytical conditions, a procedure known as sensitivity analysis [13–17]. Belshaw and Quicke [18] recently used this approach extensively in studying the evolution of a group of parasitic wasps in which species lay their eggs in either concealed or exposed hosts. They examined mapping uncertainty by modifying the parsimony cost of the evolutionary switch from exposed to concealed hosts relative to the cost of the opposite switch until their evolutionary reconstruction changed. Then they assessed phylogenetic uncertainty by finding the difference in parsimony score between the preferred tree and the best tree implying a different evolutionary history. The authors concluded that there was strong support for two switches in the parsimony reconstruction: one to exposed hosts and another in the opposite direction. Other sensitivity-type approaches to phylogenetic uncertainty include enumerating all possible phylogenetic trees consistent with some classification of the studied organisms [19] or simulating alternative trees according to some plausible tree-generating process, such as random speciation and extinction [20], and then mapping the studied character(s) onto each of these trees.

Although a particular sensitivity analysis can be illuminating, the procedure is a little bit like measuring the stability of sand castles by pouring water, blowing and stepping on them. Without standardization, the fact that one castle stands and the other falls might be solely because of differences in the treatment. Parametric statistical methods, such as likelihood analysis and Bayesian inference, can potentially add the rigor that sensitivity analysis lacks.

Mapping uncertainty

Likelihood analysis

The most common parametric approach to mapping uncertainty is based on LIKELIHOODS calculated from an evolutionary probability model. The model of choice for discrete characters is the Markov model (Box 3), exemplified by the Jukes-Cantor model and similar models long used by molecular evolutionists to study nucleotide, amino acid and codon evolution [21]. There are also Markov models for discrete characters with an arbitrary number of states [11,22–25]. For quantitative characters, Brownian motion is a popular evolutionary model [12,26]. Both Markov and Brownian motion models are mathematically convenient, but they are also able to capture many of the known complexities of the evolutionary process.

Given a fixed tree with fixed branch lengths, fixed states for the tips, and a Markov model with fixed parameters (the fixed values of which are often derived from a MAXIMUM LIKELIHOOD analysis), we can calculate the RELATIVE (CONDITIONAL) PROBABILITY of each ancestral state given the observed states at the tips [27–29]. The conditional probabilities indicate some of the uncertainty in the ancestral state assignments (Figure Ib in Box 1), but the potential error in the fixed parameters is not accounted for. For instance, assume that we were mapping a binary character (0 or 1), with the unknown parameter π_1 specifying the rate of $0 \rightarrow 1$ changes measured as a fraction of the total evolutionary rate ($\pi_0 + \pi_1$; where π_0 is this rate of $0 \rightarrow 1$ changes; Box 3). For the conditional probabilities to be valid, we have to assume the value of π_1 was known with certainty although it is typically an estimate, perhaps a maximum likelihood estimate, associated with some error.

Box 2. Phylogenetic uncertainty

Assume that we are interested in the relative rate of gain (π_1) of state 1 (green) in a character with two states (purple or green). The states could, for example, represent particular behaviours, life-history traits or morphological features. When we reconstruct the evolution of this character, it is based on a phylogenetic estimate that is associated with some error. The phylogeny could, for example, be inferred from DNA data.

The phylogenetic uncertainty can be expressed as a set of plausible alternative trees (Figure Ia). These trees can either be generated by bootstrapping the DNA data or by sampling the posterior of a Bayesian phylogenetic analysis using MCMC (Markov chain Monte Carlo) techniques. The phylogenetic uncertainty can be summarized in terms of the frequencies of the different groups of species among the alternative trees (Figure Ib). For instance, this summary indicates that only 62% of the sampled trees group species 2 and 3.

If we used bootstrapping, each sampled tree would have a single estimate of π_1 associated with it. We cannot use bootstrapping to estimate the mapping uncertainty associated with the estimation of π_1 on each tree.

In a Bayesian analysis, we would formulate priors for all parameters in the model for the DNA data as well as in the model for the purple or green character, and then sample the posterior of the composite model and the data. Instead of obtaining one π_1 value for each tree, we would implicitly be sampling each possible tree for all possible π_1 values, with each combination being sampled in proportion to its posterior probability (Table I). The frequency with which we sample each cell in the table is an estimate of the joint probability of those parameter values. After we have obtained our sample of joint probabilities, we can calculate the marginal distribution (the marginal sums in the table) for any variable in the model that we are interested in. Obviously, we would focus on the marginal distribution for π_1 (purple column), but other investigators might be more interested in the trees (bottom row) or in any of the other variables that might be included in the model.

The Bayesian marginal distribution of π_1 takes both phylogenetic and mapping uncertainty into account (Figure Ic: continuous line). This can be compared with the distribution calculated on the true (or best) tree, which reflects only mapping uncertainty (Figure Ic: dashed line). In this case, based on data generated on the tree in Box 1, the phylogenetic uncertainty flattens and shifts the posterior towards the middle, making the conclusions less certain, but only slightly. Mapping uncertainty is often going to be significant when we are trying to reconstruct ancestral states for a single character with unique evolutionary characteristics.

Table I. MCMC sampling of posterior probability

		Tree				
		1	2	3	4	Sum
π_1	0.0–0.2	52	67	22	40	**181**
	0.2–0.4	113	172	88	64	**437**
	0.4–0.6	71	82	44	41	**238**
	0.6–0.8	46	42	21	14	**123**
	0.8–1.0	7	6	4	4	**21**
	Sum	**289**	**369**	**179**	**163**	**1000**

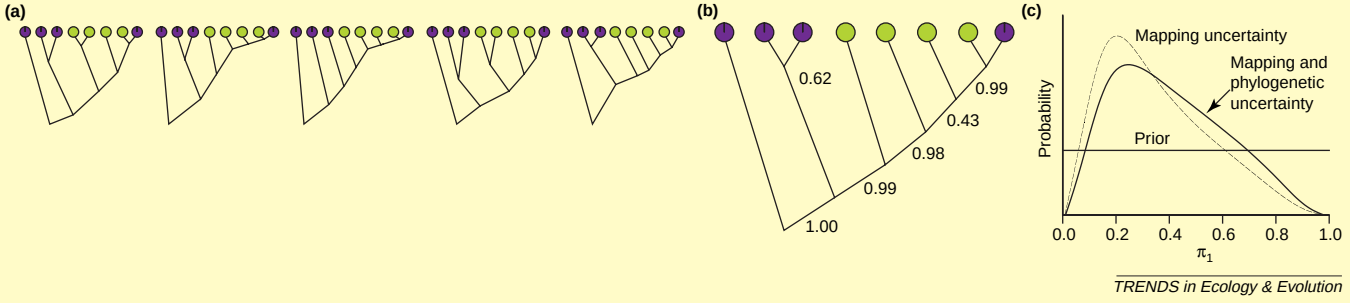

Figure I.

Bayesian inference

BAYESIAN INFERENCE provides a natural way of handling uncertainty in complex models. For the example described above, a Bayesian analysis is simply a generalization of the conditional probability analysis [30]. Instead of locking π_1 to a fixed value (Figure Ib in Box 1), a Bayesian analysis allows it to vary over the whole range of possible values (from 0 to 1) and then calculates the resulting probability, known as the POSTERIOR PROBABILITY. To do this, however, we must first determine a PRIOR PROBABILITY for all possible π_1 values because the posterior probability is proportional to the prior probability multiplied by the likelihood (Bayes' theorem). Many biologists feel uneasy about specifying priors, but it is often possible to use probability distributions that express little or no previous knowledge about the studied parameters. For instance, we can assume that all possible values of π_1 are equally likely (Figure Ic in Box 1). In calculating the probabilities of ancestral states, each value of π_1 is then weighted according to its (posterior) probability (Figure Ic in Box 1). The result is an inference of ancestral states that takes mapping uncertainty into account (Figure Id in Box 1). Bayesian inference can easily be extended to account for uncertainty in additional variables. For instance, we might wish to account for the difference in overall evolutionary rate between the mapped character and the data used to estimate the phylogeny [30–32].

Phylogenetic uncertainty

The phylogeny on which a character of interest is mapped is often based on an analysis of a large, typically molecular dataset. The first attempts to address uncertainty in phylogenetic estimates relied on BOOTSTRAPPING [33], or more precisely nonparametric bootstrapping, a method that has since become popular. In bootstrapping, large numbers of pseudoreplicate datasets are created by randomly sampling the characters in the original dataset with replacement. The chosen method of estimating phylogeny is then applied to the original character set and to each of the bootstrapped datasets. The distribution of the phylogenetic estimates from the bootstrapped data approximates the sampling distribution of the original

Box 3. Markov models

Markov models are used for random processes, in which the probability of change depends only on the current state (the Markov property). They are most easily understood in terms of their instantaneous rate matrix, which describes the transition rates in an infinitesimal amount of time. For a discrete character with two states, 0 or 1, the rate matrix Q is (Eqn I)

$$Q = \{q_{ij}\} = \begin{bmatrix} -\pi_1 & \pi_1 \\ \pi_0 & -\pi_0 \end{bmatrix}, \quad \text{[Eqn I]}$$

where q_{ij} refers to the rate in row i and column j of Q. There are two different rates in the off-diagonals: $q_{01}=\pi_1$ is the rate of $0\rightarrow1$ transitions, and $q_{10}=\pi_0$ is the rate of $1\rightarrow0$ transitions. The diagonal contains the loss rates. For instance, $q_{00}=-\pi_1$ is the rate at which the frequency of state 0 changes. The rate is negative because the frequency decreases as the character evolves from 0 to 1. The rate at which the frequency of a state decreases must balance the rate at which it evolves into other states; thus, each row in Q sums to 0.

Markov models usually tend towards an equilibrium condition (stationarity). The probability of being in a particular state i at stationarity (the stationary frequency of the state) is usually denoted π_i and can be determined from the rate matrix Q. In the binary model, the stationary frequencies correspond to the transition rates (scaling disregarded).

To illustrate this, I ran three simulations under a two-state Markov model with $\pi_1=0.75$ (and $\pi_0=0.25$). Each simulation had 200 independently evolving characters; one was started with all characters in state 0, one with all characters in state 1, and the last one with half the characters in each state. In all cases, ~75% of the characters ended up in state 1 and ~25% in state 0, as predicted by the stationary frequencies (Figure I).

To use a Markov model for simulations or probability calculations, we want to know the transition probabilities over a certain time period, t. These are represented in a matrix denoted $P(t)$, which is obtained by integrating Q over time. For the binary Markov model, we get (Eqn II)

$$P(t) = \{p_{ij}(t)\} - = \begin{bmatrix} \pi_0 + \pi_1 e^{-\mu t} & \pi_1 - \pi_1 e^{-\mu t} \\ \pi_0 - \pi_0 e^{-\mu t} & \pi_1 + \pi_0 e^{-\mu t} \end{bmatrix}, \quad \text{[Eqn II]}$$

where μ is a scaling factor. Each element of the P matrix summarizes the probability of a particular state change over an infinite number of change histories. For instance, $p_{01}(t)$ is the sum of the probability of one change $0\rightarrow1$, three changes $0\rightarrow1\rightarrow0\rightarrow1$, five changes $0\rightarrow1\rightarrow0\rightarrow1\rightarrow0\rightarrow1$, and so on, over time t.

The P matrix can be used to simulate the states at the terminals of an evolutionary tree. We draw a starting state at the root of the tree from the stationary frequencies. Then we use the P matrix for each branch in turn to generate the end state of that branch.

To obtain a sample of change histories, we need to go back to the Q matrix and utilize the fact that the waiting time (x) to the next change is exponentially distributed (Eqn III):

$$\Pr(x) = \frac{e^{-x\,/\,-q_{ii}}}{-q_{ii}}, \quad \text{[Eqn III]}$$

where $\Pr(x)$ is the probability of x and q_{ii} is the loss rate of the current state. Thus, for a binary character in state 1, the waiting time to the next change is distributed as (Eqn IV)

$$\Pr(x) = \frac{e^{-x/\pi_0}}{\pi_0}. \quad \text{[Eqn IV]}$$

When there are more than two states, the probability of the change being from i to a particular state j is determined by (Eqn V)

$$\Pr(j) = \frac{q_{ij}}{\sum_{j:j\neq i} q_{ij}}. \quad \text{[Eqn V]}$$

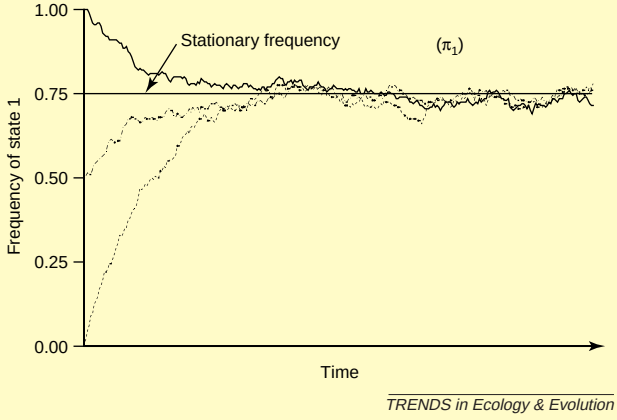

Figure I.

estimate. One of the advantages of bootstrapping is that it can be applied to a wide array of methods for reconstructing phylogeny, including distance methods, parsimony, maximum likelihood and even Bayesian inference.

Using bootstrapping to account for phylogenetic uncertainty in studies of character evolution is straightforward: simply map characters onto each of the bootstrap estimates of phylogeny and then use the distribution of these mappings to describe the effect of the uncertainty. The approach seems to have been first discussed by Felsenstein [12] and first applied by Richman and Price [34]. Ronquist and Liljeblad [35] recently used the technique extensively in reconstructing the origin of gall wasps. Simple parsimony reconstruction suggested that the first gall wasps lived in the Mediterranean and induced single-chambered galls that were distinct swellings of the seed-capsules of herbaceous Papaveraceae. When phylogenetic uncertainty was taken into account, it turned out that the robustness of these conclusions varied considerably.

Unfortunately, bootstrapping cannot be used to address mapping uncertainty. Bootstrapping DNA sequences, for example, will be of little help in understanding how precise the mapping of a single behavioural character might be on each of the possible trees. Bayesian inference, however, can account for both mapping and phylogenetic uncertainty across a heterogeneous dataset. In principle, we only need to expand the probability model to include topology, branch lengths, and other parameters necessary to infer phylogeny from the available data. The posterior probability distribution can no longer be calculated analytically (with pen and paper) because it is so complex, but we can sample from it using stochastic simulation in, for example, so-called MARKOV CHAIN MONTE CARLO (MCMC) techniques [36–40]. If the simulation is run long enough, we obtain a valid sample of the posterior probability distribution. Box 2 gives an example of how phylogenetic and mapping uncertainty are handled in a Bayesian analysis of a discrete character. Huelsenbeck and colleagues [31] first developed this approach, illustrating it with an analysis of the origin of soldier aphids. Parsimony suggested one origin and three losses of the soldier caste, but the Bayesian analysis revealed that this conclusion was uncertain. About a year later, Lutzoni and Pagel

published similar work [41,42] on the origin of lichenized fungi, but failed to apply a strict Bayesian approach to mapping uncertainty. Recently, Huelsenbeck and Rannala extended Bayesian MCMC techniques to the comparative analysis of quantitative characters using Brownian motion models [43].

An old argument is whether or not the character to be mapped should be included in the phylogenetic analysis [44–48]. The Bayesian posterior probabilities are always based on all the available data. Assume that we are mapping a behavioural trait onto a DNA phylogeny using a composite model that describes both the evolution of the DNA sequences and the behavioural character. Think of the model as a table with many dimensions, each dimension corresponding to a different parameter and each cell to a combination of parameter values (Table I in Box 2). The Bayesian MCMC analysis uses the data, the model and the prior to estimate the probability of each cell in the table (the joint probability distribution). The joint probabilities are the same regardless of whether the DNA phylogeny is derived first and the behavioural character mapped on afterwards or if both data sources are combined in a single analysis. After the analysis, the investigator is free to focus on any parameter (axis of the table) of interest by calculating its marginal distribution (the marginal sums of that dimension in the table).

Character change histories

Bayesian inference can also be used to obtain samples of character change histories from the posterior distribution while accounting for both phylogenetic and mapping uncertainty. Normally, dealing with change histories is a nuisance because there are infinitely many of them. Standard probability calculations avoid the problem by using the transition probability (P) matrix, which sums (integrates) over all possible realizations of character change; only the starting and ending states matter (Box 3). In a seminal paper, Nielsen [49] described how we can nevertheless sample change histories. The idea is to simulate character change on a set of MCMC samples, working backwards. First, we use P matrices to draw a sample of ancestral states given the observed tip states and the parameter values of the MCMC sample. We then simulate the substitution process, one branch at a time, until we get a realization that is consistent with the fixed starting and ending states of each branch (Box 4).

Nielsen's method can be used to generate a range of plausible scenarios for how evolution might have occurred. We can also study the nature of character evolution by comparing the sample of change histories leading to the observed tip states with the change histories expected from the model used for mapping [32,49]. The expected histories are obtained by simulating character evolution on the MCMC samples without fixing ancestral and tip states first; this is referred to as a posterior predictive distribution because it predicts future observations from the posterior. If the observed and expected change histories differ, we can reject the mapping model and learn something about character evolution. For instance, we might find that there is more rate variation than expected under an equal-rates model [49], we can detect positively selected sites using a model with no across-site variation in selection pressure [49], and we can reveal character correlation using a model assuming no correlation [32].

In each of these cases, it would have been easy to use a more sophisticated mapping model, enabling us to obtain a valid sample of change histories and to estimate parameters such as the extent of across-site rate variation. However, Nielsen's posterior predictive approach enables simple models to be used in addressing evolutionary phenomena that would otherwise have been difficult to model. For instance, evolutionary models that can vary across organism lineages are complicated (but not impossible) to analyze with Bayesian MCMC techniques. With Nielsen's approach, we can study the basic properties of complicated processes using simple standard models and use these results in designing more realistic models.

Bayesian controversies

Bayesian posterior probabilities have an intuitive interpretation. A tree with a posterior probability of 0.90 has a 90% chance of being true given that the model and the priors are correct. This follows from the definition of posterior probabilities and needs no mathematical proof. Nevertheless, there have been simulation studies reporting a slight Bayesian bias (underestimate) under these conditions [50–53]. This could be because of programming error, but recent analyses suggest that the bias is caused

Box 4. Generating a Bayesian sample of character change histories

To illustrate the uncertainty in reconstructing character evolution, it is useful to have a sample of likely character change histories. In the Bayesian approach, we first obtain a sample from the posterior distribution of a phylogenetic analysis (as in Figure Ia in Box 2). For each sampled tree, we draw a sample of ancestral states for the character(s) we want to map (Figure Ia). This is done by pulling conditional probabilities down the tree to obtain downpass probabilities, and then drawing ancestral states one node at a time up the tree from the downpass probabilities adjusted according to the already drawn states [49]. Once we have a sample of ancestral states, we simulate a character change history by drawing waiting times between changes one branch at a time until the drawn history matches the starting and ending states of that branch. This produces a valid sample of character change history from the posterior probability distribution (Figure Ib). A branch can have more than one change, as illustrated by the left descendant of A. By repeating the procedure for each tree in the sample, we can obtain thousands of samples of character change similar to the one in Figure Ib. These samples help reveal how the mapped characters evolve.

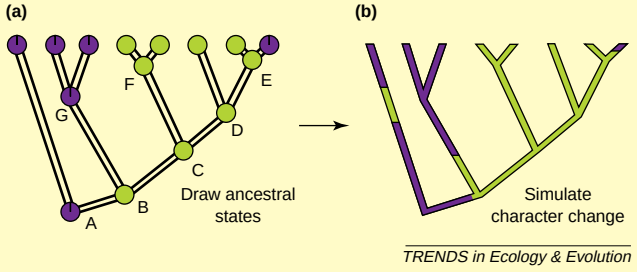

Figure I.

instead by branch lengths that are atypical (have low probability) according to the prior.

However, many workers have voiced concerns recently about the interpretation of Bayesian posterior probabilities when the model is incorrect or unknown [54–58]. When the model used for Bayesian inference is oversimplified, simulations demonstrate that erroneous trees might have high posterior probabilities [54,55]. Empirical observations also suggest that simple evolutionary models tend to be associated with less uncertainty than are more complex models [59], indicating that posterior probabilities might be too high for real data when simple models are used.

When model violation is caused by across-site heterogeneity, such as rate variation across sites, the nonparametric bootstrap can sometimes reduce the support for incorrect phylogenetic estimates [51,55]. Unfortunately, the bootstrap is unreliable under these conditions because of the model violations, so results with high bootstrap support cannot be trusted [60]. A better option, which is often available, is to use more realistic models in the Bayesian phylogenetic analysis [59,61]. Some types of model violations, including character correlation and across-tree heterogeneity, will cause overconfidence in both bootstrapped and Bayesian results [62]. The only way forward in these cases is to use better probability models for phylogenetic inference. Indeed, one of the most exciting aspects of the Bayesian MCMC approach is its efficiency in handling complex, realistic evolutionary models, including character correlation models that are not amenable to bootstrapping [62]. This is currently spurring rapid model development, which will eventually improve the overall accuracy and reliability of evolutionary inference.

When it is possible to find an adequate phylogenetic model, one could argue that the Bayesian approach is superior to bootstrapping in several ways. For example, because of the complex geometry of tree space, the bootstrap proportion is a biased estimate of a frequentist p value even when the model is correct [60,63]. Adjusting for this bias is computationally complex and is not commonly done, but a recent study suggests that uncorrected bootstrap values drastically underestimate the true values in large phylogenetic analyses [64].

Reducing uncertainty

Adding uncertainty to evolutionary inference might seem like a mixed blessing, but the Bayesian approach can also help evolutionary biologists to understand the sources of uncertainty and how to improve their reconstructions. For instance, Bayesian techniques enable us to untangle phylogenetic and mapping uncertainty (Box 2). When phylogenetic uncertainty is dominant, it is a good idea to collect more data informative about the tree. However, this is a futile solution when the major source of uncertainty is the imprecision of character mapping. In this case, one can sometimes improve the analysis considerably by adding more taxa: the larger and denser the sample of tips, the better the chances of reducing mapping uncertainty. Another possibility is to incorporate more background information when formulating the priors for the mapping model. For instance, if we were tracing the infection pathway of a virus and had reasonable previous information from other studies on its transmission rate, we could increase the precision of the evolutionary reconstruction by including this information in the analysis. This is not subjective inference worthy of suspicion; it is a smart and cost-effective way of drawing on the available information. Why not use the wheel instead of reinventing it every time a new dataset is analyzed?

Summary

With the advent of Bayesian methods, it is relatively easy to account for both mapping and phylogenetic uncertainty in reconstructing ancestral states and histories of character change for the first time. The techniques have been demonstrated for the simplest and most versatile evolutionary models, but many exciting models remain to be explored. As always, the development of user-friendly software lags behind, but free programs for Bayesian estimation of ancestral states (MRBAYES, all common platforms; http://mrbayes.net) and history of character change (SIMMAP, Mac OS X only; http://www.simmap.com) are available and more should be added in the near future. It is the beginning of an exciting new era in evolutionary reconstruction.

Acknowledgements

I am grateful to Dave Swofford, Peter Beerli, Thomas Hansen, Mike Steel and four anonymous reviewers for comments about the article.

References

1 Gaschen, B. *et al*. (2002) Diversity considerations in HIV-1 vaccine selection. *Science* 296, 2354–2360
2 Thornton, J.W. *et al*. (2003) Resurrecting the ancestral steroid receptor: ancient origin of estrogen signalling. *Science* 301, 1714–1717
3 Gabaldón, T. and Huynen, M.A. (2003) Reconstruction of the proto-mitochondrial metabolism. *Science* 301, 609
4 Schwarz, M.P. *et al*. (2003) Molecular phylogenetics of allodapine bees, with implications for the evolution of sociality and progressive rearing. *Syst. Biol.* 52, 1–14
5 Sanmartín, I. and Ronquist, F. (2004) Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Syst. Biol.* 53, 216–243
6 Waters, J.M. and Roy, M.S. (2004) Out of Africa: the slow train to Australasia. *Syst. Biol.* 53, 18–24
7 Nepokroeff, M. *et al*. (2003) Reconstructing ancestral patterns of colonization and dispersal in the Hawaiian understorey tree genus *Psychotria* (Rubiaceae): a comparison of parsimony and likelihood approaches. *Syst. Biol.* 52, 820–838
8 Johnson, M.E. *et al*. (2001) Positive selection of a gene family during the emergence of humans and African apes. *Nature* 413, 514–519
9 Chinese SARS Molecular Epidemiology Consortium. (2004) Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* 303, 1666–1669
10 Weiblen, G. (2004) Correlated evolution in fig pollination. *Syst. Biol.* 53, 128–139
11 Frumhoff, P.C. and Reeve, H.K. (1994) Using phylogenies to test hypotheses of adaptation: a critique of some current proposals. *Evolution* 48, 172–180
12 Felsenstein, J. (1988) Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* 19, 445–471
13 Swofford, D.L. and Maddison, W.P. (1992) Parsimony, character-state reconstructions, and evolutionary inferences. In *Systematics, Historical Ecology, and North American Fresh Water Fishes* (Mayden, R.L. ed), pp. 186–223, Stanford University Press

14 Maddison, W.P. and Maddison, D.R. (1992) *MacClade Version 3: Analysis of Phylogeny and Character Evolution*, Sinauer
15 Wheeler, W.C. (1995) Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Syst. Biol.* 44, 321–331
16 Donoghue, M.J. and Ackerly, D.D. (1996) Phylogenetic uncertainties and sensitivity analyses in comparative biology. *Philos. Trans. R. Soc. London B* 351, 1241–1249
17 Cunningham, C.W. *et al*. (1998) Reconstructing ancestral character states: a critical appraisal. *Trends Ecol. Evol.* 13, 361–366
18 Belshaw, R. and Quicke, D.L.J. (2002) Robustness of ancestral state estimates: evolution of life history strategy in ichneumonoid parasitoids. *Evolution* 51, 450–477
19 Losos, J.B. (1994) An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. *Syst. Biol.* 43, 117–123
20 Martins, E.P. (1996) Conducting phylogenetic comparative studies when the phylogeny is not known. *Evolution* 50, 12–22
21 Felsenstein, J. (2004) *Inferring Phylogenies*, Sinauer
22 Pagel, M. (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. London B. Biol. Sci.* 255, 37–45
23 Schultz, T.R. *et al*. (1996) The reconstruction of ancestral character states. *Evolution* 50, 504–511
24 Schluter, D. *et al*. (1995) Uncertainty in ancient phylogenies. *Nature* 377, 108–110
25 Lewis, P.O. (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50, 913–925
26 Felsenstein, J. (1985) Phylogenies and the comparative method. *Am. Nat.* 125, 1–15
27 Schluter, D. *et al*. (1997) Likelihood of ancestor states in adaptive radiation. *Evolution* 51, 1699–1711
28 Pagel, M. (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* 48, 612–622
29 Mooers, A.Ø. and Schluter, D. (1999) Reconstructing ancestor states with maximum likelihood: support for one- and two-rate models. *Syst. Biol.* 48, 623–633
30 Schultz, T.R. and Churchill, G.A. (1999) The role of subjectivity in reconstructing ancestral character states: a Bayesian approach to unknown rates, states, and transformation asymmetries. *Syst. Biol.* 48, 651–664
31 Huelsenbeck, J.P. *et al*. (2000) Accommodating phylogenetic uncertainty in evolutionary studies. *Science* 288, 2349–2350
32 Huelsenbeck, J.P. *et al*. (2003) Stochastic mapping of morphological characters. *Syst. Biol.* 52, 131–158
33 Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791
34 Richman, A.D. and Price, T. (1992) Evolution of ecological differences in the old world leaf warblers. *Nature* 355, 817–821
35 Ronquist, F. and Liljeblad, J. (2001) Evolution of the gall wasp-host plant association. *Evolution* 55, 2503–2522
36 Gamerman, D. (1997) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall
37 Lewis, P.O. (2001) Phylogenetic systematics turns over a new leaf. *Trends Ecol. Evol.* 16, 30–37
38 Huelsenbeck, J.P. *et al*. (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294, 2310–2314
39 Huelsenbeck, J.P. *et al*. (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51, 673–688
40 Holder, M. and Lewis, P.O. (2003) Phylogenetic estimation: traditional and Bayesian approaches. *Nat. Rev. Genet.* 4, 275–284
41 Lutzoni, F. *et al*. (2001) Major fungal lineages are derived from lichen symbiotic ancestors. *Nature* 411, 937–940
42 Pagel, M. and Lutzoni, F. (2002) Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation. In *Biological Evolution and Statistical Physics* (Lässig, M. and Valleriani, A. eds), pp. 151–164, Springer
43 Huelsenbeck, J.P. and Rannala, B. (2003) Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution* 57, 1237–1247
44 Coddington, J. (1988) Cladistic tests of adaptational hypotheses. *Cladistics* 4, 3–22
45 Brooks, D.R. and McClennan, D.A. (1990) *Phylogeny, Ecology, and Behavior*, University of Chicago Press
46 Armbruster, S.R. (1992) Phylogeny and the evolution of plant-animal interactions. *BioScience* 42, 12–20
47 de Queiroz, K. (1996) Including the characters of interest during tree reconstruction and the problem of circularity and bias in studies of character evolution. *Am. Nat.* 148, 700–708
48 Luckow, M. and Bruneau, A. (1997) Circularity and independence in phylogenetic tests of ecological hypotheses. *Cladistics* 13, 145–151
49 Nielsen, R. (2002) Mapping mutations on phylogenies. *Syst. Biol.* 51, 729–739
50 Wilcox, T.P. *et al*. (2002) Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phylog. Evol.* 25, 361–371
51 Douady, C.J. *et al*. (2002) Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.* 20, 248–254
52 Alfaro, M.E. *et al*. (2003) Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.* 20, 255–266
53 Cummings, M.P. *et al*. (2003) Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* 52, 477–487
54 Erixon, P. *et al*. (2003) Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* 52, 665–673
55 Suzuki, Y. *et al*. (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. U. S. A.* 99, 16138–16143
56 Buckley, T.R. (2002) Model misspecification and probabilistic tests of topology: evidence from empirical data sets. *Syst. Biol.* 51, 509–523
57 Simmons, M.P. *et al*. (2004) How meaningful are Bayesian support values? *Mol. Biol. Evol.* 21, 188–199
58 Lemmon, A.R. and Moriarty, E.C. (2004) The importance of proper model assumption in Bayesian phylogenetics. *Syst. Biol.* 53, 216–243
59 Nylander, J. *et al*. (2004) Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53, 47–67
60 Holmes, S. (2003) Bootstrapping phylogenetic trees: theory and methods. *Stat. Sci.* 18, 241–255
61 Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574
62 Galtier, N. (2004) Sampling properties of the bootstrap support in molecular phylogeny: influence of nonindependence among sites. *Syst. Biol.* 53, 38–46
63 Efron, B. *et al*. (1996) Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13429–13434
64 Sanderson, M.J. and Wojciechowski, M.F. (2000) Improved bootstrap confidence limits in large-scale phylogenies, with an example from Neo-Astragalus (Leguminosae). *Syst. Biol.* 49, 671–685