Mammal Rev. 2005, Volume 35, No. 2, 188–198. *Printed in Great Britain.*

Understanding mammalian evolution using Bayesian phylogenetic inference

CLINT D. KELLY
Department of Biology, University of Toronto at Mississauga, Mississauga, Ontario, Canada, L5L 1C6

ABSTRACT

1. Phylogenetic trees are critical in addressing evolutionary hypotheses; however, the reconstruction of a phylogeny is no easy task. This process has recently been made less arduous by using a Bayesian statistical approach. This method offers the advantage that one can determine the probability of some hypothesis (i.e. a phylogeny), conditional on the observed data (i.e. nucleotide sequences).

2. By reconstructing phylogenies using Bayes' theorem in combination with Markov chain Monte Carlo, phylogeneticists are able to test hypotheses more quickly compared with using standard methods such as neighbour-joining, maximum likelihood or parsimony. Critics of the Bayesian approach suggest that it is not a panacea, and argue that the prior probability is too subjective and the resulting posterior probability is too liberal compared with maximum likelihood.

3. These issues are currently debated in the arena of mammalian evolution. Recently, proponents and opponents of the Bayesian approach have constructed the mammalian phylogeny using different methods under different conditions and with a variety of parameters. These analyses showed the robustness (or lack of) of the Bayesian approach. In the end, consensus suggests that Bayesian methods are robust and give essentially the same answer as maximum likelihood methods but in less time.

4. Approaches based on fossils and molecules typically agree on ordinal-level relationships among mammals but not on higher-level relationships, as Bayesian analyses recognize the African radiation, Afrotheria, and the two Laurasian radiations, Laurasiatheria and Euarchontoglires, whereas fossils did not predict Afrotheria.

Keywords: Afrotheria, fossil evidence, Laurasiatheria, Markov chain Monte Carlo, phylogeny, Xenarthra

INTRODUCTION

During the last two decades, biologists have recognized the importance and utility of phylogenies in addressing many kinds of evolutionary questions (Brooks & McLennan, 1991; Huelsenbeck *et al.*, 2001). However, the phylogeny should be inferred or (re)constructed prior to incorporating a phylogeny into an analysis. The standard methods used to do this are parsimony and maximum likelihood (reviewed in Lewis, 2001). Unfortunately, these methods are often time-consuming because they 'undertake a long and arduous journey through "tree-space", a forest comprising all possible phylogenies that grows at a superexponential pace with the number of included species' (Lewis & Swofford, 2001). For example, it is not uncommon for single heuristic computer-based searches of large data sets to take days, or

Correspondence: C. D. Kelly. E-mail: cdkelly@utm.utoronto.ca

even months, to complete (Lewis, 2001). To make matters worse, likelihood approaches use bootstrapping to repeat the search hundreds of times to generate a measure of support for individual clades, thus extending an already time-intensive technique. It is not surprising then, that this method does not encourage the use of complex, and often more realistic, models of evolution (Lewis & Swofford, 2001). Even when complete, such analyses represent only a point estimate of the evolutionary history of a group of species (Lewis, 2001). Bayesian statistical methods, on the other hand, may remedy this situation by enabling quick (e.g. approximately 80% less time than maximum likelihood using bootstrapping; Leaché & Reeder, 2002) and efficient analysis of large data sets while also incorporating complex models of evolution (Huelsenbeck, Rannala & Masly, 2000b; Huelsenbeck *et al*., 2001; Lewis, 2001; Rannala, 2002). Here, I review the principles of Bayesian statistics and show how they have been applied to phylogenetic inference. I also highlight the pros and cons of using this approach vs. the more traditional method of maximum likelihood, and discuss these in the arena of mammalian evolution, a central area for this debate in evolutionary biology. I do not discuss maximum parsimony or neighbour-joining further in this paper, as they tend to figure less prominently in the debate of Bayesian performance.

In the mid-18th century, Reverend Thomas Bayes developed Bayes' rule (or theorem), the computational underpinning of Bayesian analysis. This approach uses an expanded definition of probability to assess statistical confidence, that is, probability is a direct measure of uncertainty and may or may not be a long-term frequency (Shoemaker, Painter & Weir, 1999). This is in contrast with the frequentist paradigm in which probability is defined as a long-term frequency and is viewed from the framework of hypothetically repeating an experiment many times under identical conditions (Shoemaker *et al*., 1999).

Bayesian statistics considers evidence in favour of particular parameter values, θ, which are inferred by the posterior probability distribution, $p(\theta|D)$ – the probability of some hypothesis conditional on the observed data, D. The posterior probability distribution is obtained using Bayes' theorem,

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

in which the term $p(\theta)$ represents the prior probability distribution of θ, that is, the probability distribution before examination of the data. Prior information can be based on either theoretical considerations or empirical data (Shoemaker *et al*., 1999). The probability (i.e. the likelihood) of the data given the parameters is represented by $p(D|\theta)$. The denominator $p(D)$ is the sum of the numerators $p(\theta$ and $D)$ over all possible hypotheses θ [i.e. $\Sigma_\theta\, p(\theta)\, p(D|\theta)$] and is a normalizing factor, ensuring the sum of the posterior probability is 1 (Shoemaker *et al*., 1999; Lewis, 2001). Thus, posterior probability is proportional to the product of the likelihood of the data, given the hypothesis is correct, and the prior probability of the hypothesis before any data have been collected (Lewis, 2001). That this method tells us the probability of our hypothesis given the observed data means that Bayesian results are inherently more intuitive and easier to interpret than frequentist results. For a thorough treatment of Bayesian statistical techniques, see Berger (1985) or Berry (1996).

For phylogeneticists, the salient point is that Bayesian inference and maximum likelihood differ in that the former provides probabilities of hypotheses given the data whereas the latter provides probabilities of data given hypotheses (Lewis, 2001). Thus, the application of Bayesian methods to phylogenetics yields $p(T|D)$, or the posterior probability of a tree $p(T)$ given the observed data D. However, given that a particular phylogenetic tree, T, is characterized by the topology τ and associated branch lengths β (Cummings *et al*., 2003), phylogeneticists

are often most interested in $p(\tau|D)$ or the probability of a topology conditional on the observed data.

The relatively limited usage of a Bayesian approach in evolutionary biology is likely due to a combination of it being a relatively obscure and unknown method as well as mathematically complex. However, this is no longer the case, as increasingly more research into Bayesian statistics has raised its profile in many scientific fields. More practically, the advent of powerful desktop computers now permits computationally demanding numerical techniques, such as Monte Carlo simulation, to handle complex models in Bayesian analyses (Shoemaker *et al*., 1999; Huelsenbeck *et al*., 2000b, 2001) in easy-to-use software packages (e.g. MrBayes; Huelsenbeck & Ronquist, 2001). Rannala (2002) suggests that biologists have been justified in dismissing Bayesian approaches because, historically, Bayesian inference was limited to simple models for which analytical solutions were available and models were too often chosen based on mathematical convenience. Now, however, given current computational power and numerical techniques, such as Markov chain Monte Carlo (MCMC) (Metropolis *et al*., 1953; Hastings, 1970), prior distributions and likelihoods can be selected that are more suitable to the question(s) being addressed (Rannala, 2002). Thus, multiparametered phylogenetic problems with complicated posteriors are now possible to solve (Shoemaker *et al*., 1999; Lewis, 2001; Rannala, 2002).

BAYESIAN INFERENCE AND MARKOV CHAIN MONTE CARLO

Bayesian inference of phylogeny generates an approximation of the posterior probability distribution of all parameters (i.e. tree topology, branch lengths, and substitution model parameter estimates) using MCMC (Huelsenbeck *et al*., 2000b, 2001; Lewis, 2001; Holder & Lewis, 2003; Beaumont & Rannala, 2004). Briefly, implementation of MCMC means that a Markov chain undertakes a correlated random 'walk' through parameter space while periodically sampling values according to rules provided by the investigator, typically every 100 or 1000 steps. Because values are sampled in proportion to their probability distribution, a posterior probability is approximated. Applied to phylogenetic analyses, each step in a Markov chain involves a random modification of tree topology, a branch length or a parameter in the substitution model. If the computed posterior probability is larger than the probability of the current tree topology and parameter values, then the proposed step is taken and the cycle repeated. If the proposed step has a lower posterior probability, then it is accepted under particular conditions such that small steps downward are accepted more often than large ones. If the proposed step is rejected, then the present location is added to the chain and the cycle repeated (Lewis, 2001; Holder & Lewis, 2003). As this process continues, trees are sampled in proportion to their posterior probability of occurrence under the specified model of gene-sequence evolution (e.g. time-reversible model, Swofford *et al*., 1996). Therefore, the value that one gets represents the nodal support for that particular phylogenetic tree (reviewed in Lewis, 2001). Typically, the first trees generated are removed from the sample to avoid including any trees generated before convergence of the Markov chain onto the posterior probability distribution (Huelsenbeck *et al*., 2002). Convergence is defined as the tendency for a mathematical function to approach some value, or set of values, with increasing n (Beaumont & Rannala, 2004). In MCMC parlance, n represents the number of runs, and the values that the chain approaches are the posterior probabilities (Beaumont & Rannala, 2004). The preconvergence period is called the 'burn-in period'. For example, in their comparative analysis of fungal evolution, Lutzoni, Pagel & Reeb (2001) generated 200 000 phylogenetic trees, sampled every 10th tree and then removed the first 100 trees that represented the burn-in period. Note that this burn-in period is considerably shorter than

other values reported in the literature (e.g. 30 000, Alfaro, Zoller & Lutzoni, 2003; 100 000, Springer *et al.* 2003; 25 000, Lemmon & Moriarty, 2004).

Bayesian inference is straightforward, and the results are easy to understand compared with more traditional approaches (e.g. maximum likelihood), because the posterior probability represents the support for ancestral nodes (Huelsenbeck & Rannala, 2004). If one was interested in knowing the probability of group X being monophyletic, all they would have to do is run a Markov chain and sample trees periodically. Because the Markov chain visits trees in proportion to their posterior probability, the value generated is the probability of the group being monophyletic. Therefore, if in a sample of 100 000 trees, group X is monophyletic 82 391 times, the probability of group X being monophyletic is approximately 0.82391, given the observed data.

Choice of evolutionary model and the rate of evolution

Evolutionary models describe sequence evolution or, more specifically, nucleotide substitution which is typically modelled as a stochastic process that can be either time-homogeneous, time-reversible or time-heterogeneous, time-irreversible. Choosing the correct model of evolution from the many available (see Swofford *et al*., 1996) is critical, because using inappropriate models can result in incorrect tree topologies (i.e. poor performance of phylogenetic methods) and inaccurate parameter estimation (e.g. underestimation of sequence divergence) which, consequently, lead to erroneous interpretations of the evolutionary process (reviewed in Sullivan & Swofford, 1997). This is important when using the Bayesian approach, because this method is particularly sensitive, compared with the bootstrap, to model misspecification (Huelsenbeck & Rannala, 2004). Surprisingly, many researchers simply choose the default option in their computer software for phylogenetic estimation (Posada & Crandall, 2001). The best-fit model has traditionally been chosen using maximum likelihood approaches that use either the likelihood ratio test or parametric bootstrapping. Although this approach has been shown to be adequate (Posada & Crandall, 2001), there is room for improvement because it requires that tree topology and model parameters be assumed.

Bollback (2002) argues that phylogenetic statistical methods should strive to reduce the number of assumptions in the analysis. Bayesian methods achieve this because they accommodate uncertainty in topology, branch lengths and substitution rates (Huelsenbeck *et al*., 2000b). Suchard, Weiss & Sinsheimer (2001) recently developed a method of model selection that uses Bayes factors, the ratio of the posterior probability of one model to that of another divided by the ratio of the prior probabilities. This is the Bayesian analogue of the likelihood ratio test. Although their method is an advance in model assessment techniques, it is not perfect because it does not evaluate the overall or global adequacy of a model – it simply compares two models (i.e. performs a relative assessment) (Bollback, 2002). Bollback (2002) remedied this fault by developing a Bayesian method that evaluates the absolute adequacy of the evolutionary model. An added bonus of Bollback's (2002) method is that it permits a wide range of test statistics to be calculated that examine various aspects of model performance.

A recent important advance in Bayesian phylogenetic inference is a model permitting the analysis of data sets combining morphology and molecules (Nylander *et al*., 2004). In their inference of relationships among gall wasps, Nylander *et al*. (2004) found that although morphology contributed less than 5% of the characters in the data set, it nevertheless influenced their interpretation of the phylogeny. They argue that morphological characters contribute important information and, where applicable, must not be ignored.

A controversial and active field of research involves estimating the rate of evolution using the rate of molecular evolution (Cooper & Fortey, 1998; Bromham, Phillips & Penny, 1999;

Smith & Peterson, 2002; Felsenstein, 2004). The most popular method for estimating divergence times is to use Zuckerkandl & Pauling's (1965) molecular clock, which suggests that all evolutionary lineages have experienced sequence changes at identical rates. Considerable theoretical and empirical attention has often rejected the molecular clock in favour of alternative models (Thorne, Kishino & Painter, 1998; Huelsenbeck, Larget & Swofford, 2000a; Kishino, Thorne & Bruno, 2001; Aris-Brosou & Yang, 2002; Thorne & Kishino, 2002) that relax molecular clock assumptions and thereby recognize rate heterogeneity across different nucleotide positions, different genes, different genomic regions or different genomes within an organismal lineage (Hillis, Mable & Moritz, 1996). These more realistic and relaxed clock models harness the computational power of Bayesian inference, and permit variation in rates by specifying a prior distribution of the evolutionary rates along branches and then having an MCMC algorithm assign rates to different parts of the phylogenetic tree and sampling possible topologies (Thorne *et al*., 1998; Kishino *et al*., 2001). Although the Bayesian models are tedious to apply, they are considered the most accurate available (Felsenstein, 2004), and accommodate all models of nucleotide substitution and amino acid replacement (Thorne *et al*., 1998), with some permitting constraints imposed by the fossil record (Kishino *et al*., 2001). Despite these advantages, the divergence times in mouse lemur species were 'remarkably similar' in a comparison of likelihood and Bayesian methods (Yang & Yoder, 2003).

Criticisms of the Bayesian approach

The Bayesian approach using MCMC is a significant advance in evolutionary biology, because it permits the relatively rapid analysis of large data sets with complex models; however, it is important to realize that this method is not a panacea. Because MCMC analyses are typically started at a point far removed from regions of high posterior probability, it takes some time to locate the probable parameter values and generate a reasonable posterior distribution, in other words, attain convergence (Huelsenbeck *et al*., 2002). Knowing how long to run chains in order to obtain an adequate approximation of the posterior probabilities of trees presents one of the biggest problems of this method and is currently under debate (Huelsenbeck *et al*., 2001, 2002; Lewis, 2001; Lewis & Swofford, 2001). One possible means of alleviating this problem is to run several chains, each starting from a different and randomly selected point in parameter space (Lewis, 2001; Huelsenbeck *et al*., 2002). Resulting approximations that are similar would strongly indicate convergence and that the chains were run for sufficient periods of time (Lewis, 2001; Huelsenbeck *et al*., 2002; see also Murphy *et al*., 2001b). This aspect of Bayesian inference requires more research.

Bayesian MCMC methods can handle complex evolutionary models; however, given that complex is a relative term, phylogeneticists must be aware that models currently cannot be as complex as desired (Lewis & Swofford, 2001; Rannala, 2002). Not recognizing the limitations of model complexity will lead investigators to over-parameterize the analysis (i.e. the chosen model is more complex than the true underlying model) and, thus, produce a model with unidentifiable parameters (Rannala, 2002). Despite the possibility of over-parameterization, Bayesian inference remains less prone to this problem than maximum likelihood techniques (Thorne *et al*., 1998). As a consequence of over-parameterization, different combinations of unidentifiable parameters generate identical likelihoods (the probability of the data given the parameters), thus, rendering it impossible to decide among the potential parameter values based on the data (Rannala, 2002). Moreover, over-parameterization increases the influence of priors (despite the size of the data set, see below) and, more practically, may slow the rate of convergence of the MCMC algorithm (Rannala, 2002). Given that over-parameterization is a property of the likelihood, similar problems are also likely

with maximum likelihood analyses, and should therefore not be highlighted as a problem unique to Bayesian approaches (Rannala, 2002). Despite these costs, using over-complicated models is less of a concern than using simple models because, in simulations, simple models underestimate posterior probabilities while over-parameterized models result in negligible overestimates of posterior probabilities (Huelsenbeck & Rannala, 2004). In the future, researchers will likely be less restrained by computational ability, as methods (e.g. MCMC) should improve to accommodate more complex probability models (e.g. Marjoram *et al*., 2003). Readers are encouraged to consult Rannala (2002) and Nylander *et al*. (2004) for a detailed discussion of how to deal with over-parameterization in theory and in practice, respectively.

Another common criticism of the Bayesian approach is that the choice of the prior distribution is too subjective (Shoemaker *et al*., 1999), an important point to consider because the results (the posterior distribution) can be sensitive to the prior (Huelsenbeck *et al*., 2002). For example, two researchers using the same data set could produce different results by using different priors. Some critics even argue that the subjective nature of Bayesian probabilities contradicts the goal of objectivity in science (Huelsenbeck *et al*., 2002); however, that debate is beyond the scope of this paper. In order to satisfy critics, Lewis (2001) recommends that subjectivity inherent in the prior be explicit and defensible. Huelsenbeck *et al*. (2002) are strongly in favour of incorporating prior beliefs into an analysis, because background knowledge should not be ignored.

Specifying prior distributions for all model parameters allows one to integrate out all other parameters (e.g. branch lengths, DNA substitution parameter values), while making inferences about the parameter of interest (e.g. tree topology) (Huelsenbeck *et al*., 2002). This is a statistically robust approach if one possesses strong prior beliefs; however, if one does not, then specifying prior probability distributions can be worrisome (Huelsenbeck *et al*., 2002). When this situation arises, systematists typically give all trees equal weight (Huelsenbeck *et al*., 2002). The controversy over prior distributions may be irrelevant, because as data sets become increasingly large in a Bayesian analysis, the effect of the priors on the posterior distribution decreases (Huelsenbeck *et al*., 2002).

Clade probabilities generated by bootstrap probabilities in maximum likelihood analyses are often lower than Bayesian posteriors, a condition attributed to bootstrap values being too conservative (Wilcox *et al*., 2002; Alfaro *et al*., 2003). However, the alternative hypothesis that Bayesian probabilities are too high or liberal (Cummings *et al*., 2003) seems to be consistently ignored (Suzuki, Glazko & Nei, 2002). As mentioned above, this discrepancy may be the result of model choice, and would likely be resolved by implementing appropriate evolutionary models (Huelsenbeck & Rannala, 2004). This criticism will be discussed in more detail below as it has come to bear on the debate of mammalian evolution.

BAYESIAN INFERENCE AND THE EVOLUTION OF MAMMALS

Despite being a relatively new tool in evolutionary biology, Bayesian phylogenetic methods have already demonstrated their utility in several ways, including comparative analyses, molecular clock investigations and in phylogenetic inference (reviewed in Huelsenbeck *et al*., 2001). In particular, Bayesian methods have been used extensively in elucidating the timing of the placental mammal radiation, as well as higher-level relationships among the major mammalian groups, two of the most controversial topics in evolutionary biology.

The Explosive, Long Fuse and Short Fuse models are three hypotheses proposed to explain placental mammal diversification (Archibald & Deutschman, 2001). All hypotheses consider mammalian divergence relative to the Cretaceous–Tertiary (K/T) boundary, 65 million years

ago. The Explosive model suggests most if not all ordinal originations and intraordinal diversification occurred after the K/T boundary. Similar to this model, The Long Fuse model postulates that intraordinal diversifications followed the K/T boundary; however, interordinal divergences are found well back in the Late Cretaceous. The Short Fuse model argues that interordinal divergence and some intraordinal diversification occurred more than 100 million years ago. Fossil evidence, according to Archibald & Deutschman (2001), suggests that 15 of 18 extant orders of placental mammals began to diversify sometime during the first 16 million years of the Cenozoic, thus supporting both the Long Fuse or Explosive models of radiation. Some molecular studies support the Short Fuse model of interordinal divergence during the Cretaceous period (e.g. Kumar & Hedges, 1998; Penny *et al*., 1999; see also Bromham *et al*., 1999). Recently, Springer *et al*. (2003) investigated this issue using Murphy *et al*.'s (2001b) large molecular data set (see below for details), in conjunction with the above-mentioned Bayesian approaches for estimating posterior probabilities of divergence times (Thorne *et al*., 1998; Kishino *et al*., 2001). Recall that the Thorne-Kishino relaxed clock model allows constraints from the fossil record in addition to permitting rates of molecular evolution to vary on different branches of a phylogenetic tree. Springer *et al*.'s (2003) Bayesian analysis provided robust results (i.e. with three different priors for the placental root, two different marsupial out-groups, different data partitions, fossil constraints and controls for taxon body size; see Springer *et al*., 2003 for details of analysis) that were in general agreement with the fossil record and, thus, the Long Fuse model. They found no evidence to support the Short Fuse model.

Higher-level relationships among mammal groups have also been addressed using Bayesian approaches. Two independent studies (Madsen *et al*., 2001; Murphy *et al*., 2001a), each using a maximum likelihood approach based on the concatenation (pooling data from multiple genes) of mitochondrial RNA genes and nuclear genes, resolved placental orders into four major clades: (i) Afrotheria (elephants, manatees, hyraxes, tenrecs, aardvark and elephant shrews), (ii) Xenarthra (sloths, anteaters, armadillos), (iii) Euarchontoglires (rodents, lagomorphs, flying lemurs, tree shrews, primates) and (iv) Laurasiatheria (bats, cetaceans, cats, dogs). The latter two clades have a Northern Hemisphere origin, and are considered sister taxa that together constitute a clade named Boreoeutheria (Springer & de Jong, 2001). These results provided some support for a Gondwanan origin for extant placental mammals, and highlighted the influence of tectonic events in their early differentiation. However, the relationships among the four groups were not completely resolved because of the sensitivity of the root to the phylogenetic method used (i.e. parsimony, distance methods, maximum likelihood) (Madsen *et al*., 2001). The most likely position of the root on the placental tree is either (i) the base of Afrotheria, (ii) the base of Xenarthra, or (iii) the branch separating Xenarthra and Afrotheria from Boreoeutheria. To better resolve hierarchical relationships both within the four superordinal groups and at deeper levels in the eutherian phylogenetic tree, Madsen *et al*. (2001) and Murphy *et al*. (2001a) combined and expanded their data sets and used a Bayesian approach to analyse it (Murphy *et al*., 2001b; see also, Jow *et al*., 2002).

Murphy *et al*.'s (2001b) combined molecular data set comprised 19 nuclear and three mtDNA gene sequences for 42 placentals (representing all major lineages) and two marsupial out-groups. They implemented three separate MCMC runs, each starting with random trees for each of four simultaneous chains. That the runs provided similar joint posterior probability distributions for the tree topology and the estimated parameters of the model for sequence evolution (i.e. general-time-reversible + gamma + invariants model), argue the authors, is sufficient evidence that the chains were run long enough and sampled the same posterior probability landscape (Murphy *et al*., 2001b). Nearly the entire superordinal pla-

cental tree was resolved with posterior probabilities ≥0.98; only the split between the Perissodactyia and Carnivora + Pholldota (posterior probability = 0.74) and the split between Sirenia and Hyracoidea (0.76) were lower. Because an identical tree was generated using maximum likelihood, Murphy *et al*. (2001b) were confident in their Bayesian result.

Murphy *et al*. (2001b) found that Bayesian posterior probabilities of interior branches were generally higher than non-parametric maximum likelihood bootstrap values, a result consistent with the reported conservative nature of bootstrap values (Alfaro *et al*., 2003). In a re-analysis of Murphy *et al*.'s (2001b) data, however, Suzuki *et al*. (2002) showed by computer simulation that Bayesian posteriors were 'excessively liberal' whereas bootstrap probabilities in maximum likelihood and neighbour-joining analyses were 'slightly conservative' (see also Cummings *et al*., 2003; Delsuc, Stanhope & Douzery, 2003). This result held up when either concatenated or completely linked sequences were analysed (Suzuki *et al*., 2002). Alfaro *et al*. (2003) suggest that bootstrap values and posterior probabilities each provide useful confidence measures of the data, depending on what the phylogeneticists would like their confidence method to measure. They suggest 'nonparametric bootstrapping is appropriate if one is interested in the sensitivity of observed results to the sampling error associated with collecting characters from a hypothesized underlying character distribution' (Alfaro *et al*., 2003). Conversely, Bayesian posterior probabilities are appropriate 'when one is willing to specify a fully probabilistic model of character evolution and wishes to place confidence limits on the results of analysis conditioned on the observed data and that model' (Alfaro *et al*., 2003). In light of their findings, Suzuki *et al*. (2002) recommend a re-examination of those studies in which molecular phylogenies were obtained by Bayesian phylogenetics (e.g. Karol *et al*., 2001; Lutzoni *et al*., 2001; Buckley *et al*., 2002; Leaché & Reeder, 2002).

Murphy *et al*.'s (2001b) Bayesian analysis firmly supported Laurasiatheria and Euarchontoglires as sister taxa in a clade named Boreoeutheria (see also Springer & de Jong, 2001), and placed Xenarthra and Boreoeutheria as sister taxa. This analysis supported the basal position of Afrotheria on the placental tree (Murphy *et al*., 2001b). That Jow *et al*. (2002) obtained identical results in their Bayesian phylogenetic analysis of the complete set of mitochondrial tRNA and rRNA sequences from 54 mammal species strengthens the confidence in Murphy *et al*.'s (2001b) analysis. Although fossils and molecules typically agree on ordinal-level relationships, the recognition of the African radiation, Afrotheria, and the two Laurasian radiations, Laurasiatheria and Euarchontoglires, represent substantial differences between the two methods of phylogenetic investigation; fossils did not predict Afrotheria (Archibald, 2003). This is important because if the Afrotheria and Laurasiatheria hypotheses are correct then 'morphology has failed to recover some of the most fundamental clades in all of Eutheria' (Madsen *et al*., 2001, p. 613).

The molecular phylogenetic analyses of mammalian evolution (Madsen *et al*., 2001; Murphy *et al*., 2001a,b; Jow *et al*., 2002) completed to date, may have suffered because Xenarthra were poorly represented taxonomically (only three taxa) relative to the other three clades (8 afrotherian, 11 euarchonotglirian, 20 laurasiatherian) (Delsuc *et al*., 2002). It is possible that with more extensive sampling within Xenarthra, the placement of the root on the placental tree could be better resolved (Delsuc *et al*., 2002). Delsuc *et al*. (2002) addressed this problem, by increasing the sample of xenarthran species from three (e.g. Murphy *et al*., 2001b) to 13, thus representing all but one of the living genera. Their Bayesian analysis based on three nuclear genes was fully compatible with previous molecular analyses (Madsen *et al*., 2001; Murphy *et al*., 2001a;b; Jow *et al*., 2002) and supported the four major clades (Afrotheria, Xenarthra, Laurasiatheria and Euarchontoglires) and the grouping of the North American clades (Laurasiatheria and Euarchontoglires) into Boreoeutheria (Murphy *et al*., 2001b;

Springer & de Jong, 2001). Again, suggesting that Murphy *et al*.'s (2001b) Bayesian inference of eutherian phylogeny was robust. A noteworthy aspect of the Murphy *et al*.'s study was that the data set was larger (i.e. number of base pairs and genes used) than previous studies. Such a data set is less prone to sampling error and would perhaps give similar answers regardless of the method of analysis used.

In summary, the three studies (Murphy *et al*., 2001b; Delsuc *et al*., 2002; Jow *et al*., 2002) that addressed mammalian evolution using Bayesian methods produced similar results and, in each study, maximum likelihood supported the most probable Bayesian topology (but see Suzuki *et al*., 2002). Their independent analyses provide strong evidence for the most recent common ancestry of placental mammals in Gondwana, contradicting fossil evidence of a Laurasian origin (Archibald, 2003). Investigators must be aware that discrepancies between molecular and morphological approaches to measuring divergence and diversification arise, because they are effectively measuring different aspects of phylogeny (Smith & Peterson, 2002). That is, molecular methods measure time of separation (genotypic level of analysis), whereas fossil data estimate time of differentiation (phenotypic level of analysis) with the true time of divergence falling closer to the molecular estimate (reviewed in Smith & Peterson, 2002).

Although a field in its infancy and still obviously feeling growing pains, Bayesian phylogenetics will only become a more and more powerful tool in evolutionary biology. As technology improves and mathematical coprocessors achieve higher capacities, systematists will be capable of running more complex and realistic models of evolution in less time than is currently required. The advantages of this approach (e.g. reduced time required for analysis, analyses with complex and realistic models of evolution are tractable) seem to far outweigh the criticisms which will surely become negligible as time passes and phylogeneticists devote more time and effort to their resolution. Perhaps in the near future, the entire tree of life will be resolved using Bayesian inference.

ACKNOWLEDGEMENTS

I thank Darryl Gwynne, Bruce Rannala and two anonymous referees for comments on the manuscript.

REFERENCES

Alfaro, M.E., Zoller, S. & Lutzoni, F. (2003) Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution*, **20**, 255–266.

Archibald, J.D. (2003) Timing and biogeography of the eutherian radiation: fossils and molecules compared. *Molecular Phylogenetics and Evolution*, **28**, 350–359.

Archibald, J.D. & Deutschman, D.H. (2001) Quantitative analysis of the timing of the origin and diversification of extant placental orders. *Journal of Mammalian Evolution*, **8**, 107–124.

Aris-Brosou, S. & Yang, Z.H. (2002) Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA Phylogeny. *Systematic Biology*, **51**, 703–714.

Beaumont, M.A. & Rannala, B. (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*, **5**, 251–261.

Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. Springer-Verlag, New York.

Berry, D.A. (1996) *Statistics: A Bayesian Perspective*. Wadsworth Publishing, Duxbury Press, Belmont, CA.

Bollback, J.P. (2002) Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution*, **19**, 1171–1180.

Bromham, L., Phillips, M.J. & Penny, D. (1999) Growing up with dinosaurs: molecular dates and the mammalian radiation. *Trends in Ecology and Evolution*, **14**, 113–118.

Brooks, D.R. & Mclennan, D.A. (1991) *Phylogeny, Ecology, and Behavior: A Research Program in Comparative Biology*. The University of Chicago Press, Chicago.

Buckley, T.R., Arensburger, P., Simon, C. & Chambers, G.K. (2002) Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. *Systematic Biology*, **51**, 4–18.

Cooper, A. & Fortey, R. (1998) Evolutionary explosions and the phylogenetic fuse. *Trends in Ecology and Evolution*, **13**, 151–156.

Cummings, M.P., Handley, S.A., Myers, D.S., Reed, D.L., Rokas, A. & Winka, K. (2003) Comparing bootstrap and posterior probability values in the four-taxon case. *Systematic Biology*, **52**, 477–487.

Delsuc, F., Scally, M., Madsen, O., Stanhope, M.J., De Jong, W.W., Catzeflis, F.M., Springer, M.S. & Douzery, E.J.P. (2002) Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Molecular Biology and Evolution*, **19**, 1656–1671.

Delsuc, F., Stanhope, M.J. & Douzery, E.J.P. (2003) Molecular systematics of armadillos (Xenarthra, Dasypodidae): contribution of maximum likelihood and Bayesian analyses of mitochondrial and nuclear genes. *Molecular Phylogenetics and Evolution*, **28**, 261–275.

Felsenstein, J. (2004) *Inferring Phylogenies*. Sinauer associates, Inc., Sunderland, MA.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Hillis, D.M., Mable, B.K. & Moritz, C. (1996) Applications of molecular systematics: the state of the field and a look to the future. In: *Molecular Systematics* (Ed. by D.M. Hillis, C. Moritz & B.K. Mable), pp. 515–543. Sinauer Associates, Inc., Sunderland, MA.

Holder, M. & Lewis, P.O. (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics*, **4**, 275–284.

Huelsenbeck, J.P., Larget, B., Miller, R.E. & Ronquist, F. (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology*, **51**, 673–688.

Huelsenbeck, J.P., Larget, B. & Swofford, D. (2000a) A compound Poisson process for relaxing the molecular clock. *Genetics*, **154**, 1879–1892.

Huelsenbeck, J.P. & Rannala, B. (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology*, **53**, 904–913.

Huelsenbeck, J.P., Rannala, B. & Masly, J.P. (2000b) Accommodating phylogenetic uncertainty in evolutionary studies. *Science*, **288**, 2349–2350.

Huelsenbeck, J.P. & Ronquist, F. (2001) MrBayes: Bayesian inference of phylogeny. *Biometrics*, **17**, 754–755.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R. & Bollback, J.P. (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, **294**, 2310–2314.

Jow, H., Hudelot, C., Rattray, M. & Higgs, P.G. (2002) Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. *Molecular Biology and Evolution*, **19**, 1591–1601.

Karol, K.G., Mccourt, R.M., Cimino, M.T. & Delwiche, C.F. (2001) The closest living relatives of land plants. *Science*, **294**, 2351–2353.

Kishino, H., Thorne, J.L. & Bruno, W.J. (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution*, **18**, 352–361.

Kumar, S. & Hedges, S.B. (1998) A molecular timescale for vertebrate evolution. *Nature*, **392**, 917–920.

Leaché, A.D. & Reeder, T.W. (2002) Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. *Systematic Biology*, **51**, 44–68.

Lemmon, A.R. & Moriarty, E.C. (2004) The importance of proper model assumption in Bayesian phylogenetics. *Systematic Biology*, **53**, 265–277.

Lewis, P.O. (2001) Phylogenetic systematics turns over a new leaf. *Trends in Ecology and Evolution*, **16**, 30–37.

Lewis, P.O. & Swofford, D.L. (2001) Back to the future: Bayesian inference arrives in phylogenetics. *Trends in Ecology and Evolution*, **16**, 600–601.

Lutzoni, F., Pagel, M. & Reeb, V. (2001) Major fungal lineages are derived from lichen symbiotic ancestors. *Nature*, **411**, 937–940.

Madsen, O., Scally, M., Douady, C.J., Kao, D.J., Debry, R.W., Adkins, R., Amrine, H.M., Stanhope, M.J., De Jong, W.W. & Springer, M.S. (2001) Parallel adaptive radiations in two major clades of placental mammals. *Nature*, **409**, 610–614.

Marjoram, P., Molitor, J., Plagnol, V. & Tavaré, S. (2003) Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the USA*, **100**, 15324–15328.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E. (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A. & O'brien, S.J. (2001a) Molecular phylogenetics and the origins of placental mammals. *Nature*, **409**, 614–618.

Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., De Jong, W.W. & Springer, M.S. (2001b) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, **294**, 2348–2351.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P. & Nieves-Aldrey, J.L. (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology*, **53**, 47–67.

Penny, D., Hasegawa, M., Waddell, P.J. & Hendy, M.D. (1999) Mammalian evolution: timing and implications from using the LogDeterminant transform for proteins of differing amino acid composition. *Systematic Biology*, **48**, 76–93.

Posada, D. & Crandall, K.A. (2001) Selecting the best-fit model of nucleotide substitution. *Systematic Biology*, **50**, 580–601.

Rannala, B. (2002) Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Systematic Biology*, **51**, 754–760.

Shoemaker, J.S., Painter, I.S. & Weir, B.S. (1999) Bayesian statistics in genetics: a guide for the uninitiated. *Trends in Genetics*, **15**, 354–358.

Smith, A.B. & Peterson, K.J. (2002) Dating the time of origin of major clades: molecular clocks and the fossil record. *Annual Review of Earth and Planetary Sciences*, **30**, 65–88.

Springer, M.S. & De Jong, W.W. (2001) Which mammalian supertree to bark up? *Science*, **291**, 1709–1711.

Springer, M.S., Murphy, W.J., Eizirik, E. & O'Brien, S.J. (2003) Placental mammal diversification and the Cretaceous–Tertiary boundary. *Proceedings of the National Academy of Sciences of the USA*, **100**, 1056–1061.

Suchard, M.A., Weiss, R.E. & Sinsheimer, J.S. (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution*, **18**, 1001–1013.

Sullivan, J. & Swofford, D.L. (1997) Are guinea pigs rodents? The importance of adequate models in molecular phylogenies. *Journal of Mammalian Evolution*, **4**, 77–86.

Suzuki, Y., Glazko, G.V. & Nei, M. (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences of the USA*, **99**, 16138–16143.

Swofford, D.L., Olsen, G.J., Waddell, P.J. & Hillis, D.M. (1996) Phylogenetic inference. In: *Molecular Systematics* (Ed. by D.M. Hillis, C. Moritz & B.K. Mable), pp. 407–514. Sinauer Associates, Inc., Sunderland, MA.

Thorne, J.L. & Kishino, H. (2002) Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology*, **51**, 689–702.

Thorne, J.L., Kishino, H. & Painter, I.S. (1998) Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, **15**, 1647–1657.

Wilcox, T.P., Zwickl, D.J., Heath, T.A. & Hillis, D.M. (2002) Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Molecular Phylogenetics and Evolution*, **25**, 361–371.

Yang, Z.H. & Yoder, A.D. (2003) Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Systematic Biology*, **52**, 705–716.

Zuckerkandl, E. & Pauling, L. (1965) Evolutionary divergence and convergence in proteins. In: *Evolving Genes and Proteins* (Ed. by V. Bryson & H.J. Vogel), pp. 97–166. Academic Press, New York.

Submitted 2 February 2004; returned for revision 22 March 2004; revision accepted 30 November 2004
Editor: RM