Complecto errorem: Embracing Uncertainty in Molecular Phylogenetic Analysis

by

Joseph W. Brown

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in The University of Michigan
2010

Doctoral Committee:

Associate Professor Yin-Long Qiu, Co-Chair
Professor David P. Mindell, California Academy of Sciences, Co-Chair
Assistant Professor Inés Ibáñez
Assistant Professor Aaron A. King

If we view statistics as a discipline in the service of science, and science as being an attempt to understand (i.e., model) the world around us, then the ability to reveal sensitivity of conclusions from fixed data to various model specifications, all of which are scientifically acceptable, is equivalent to the ability to reveal boundaries of scientific uncertainty. When sharp conclusions are not possible without obtaining more information, whether it be more data, new theory, or deeper understanding of existing data and theory, then it must be scientifically valuable and appropriate to expose this sensitivity and thereby direct efforts to seek the particular information needed to sharpen conclusions.

Donald B. Rubin (1984) *The Annals of Statistics* 12: 1151-1172

We must confine ourselves to those forms which we know how to handle, or for which any tables which may be necessary have been constructed. More or less elaborate form will be suitable according to the volume of the data.

R. A. Fisher (1922) *Philosophical Transactions of the Royal Society* 222: 309-368

'...And then comes the grandest idea of all! We made a map of the country, on the scale of a mile to the mile!'
'Have you used it much?' I enquired.
'It has never been spread out yet,' said Mein Herr, 'The farmers objected: they said it would cover the whole country and shut out the sunlight! So we now use the country itself, as its own map, and I assure you it does nearly as well.'

Lewis Carroll (1893) Sylvie and Bruno Concluded

I went on to test the program in every way I could devise. I strained it to expose its weaknesses. I ran it for high-mass stars and low-mass stars, for stars born exceedingly hot and those born relatively cold. I ran it assuming the superfluid currents beneath the crust to be absent -- not because I wanted to know the answer, but because I had developed an intuitive feel for the answer in this particular case. Finally I got a run in which the computer showed the pulsar's temperature to be less than absolute zero. I had found an error. I chased down the error and fixed it. Now I had improved the program to the point where it would not run at all.

George Greenstein (1984) Frozen Star: Of Pulsars, Black Holes and the Fate of Stars

I love deadlines. I like the whooshing sound they make as they fly by.

Douglas Adams

2010

To My Wife, Rachel

Acknowledgements

Many people facilitated directly or indirectly the completion of this work.

I am grateful for my dissertation committee, both past and present, for encouragement throughout: Thomas Duda, Philip Gingerich, Inés Ibáñez, Aaron King, David Mindell, Robert Payne, and Yin-Long Qiu. Tom, in particular, deserves special thanks for agreeing to serve as my on-site supervisor for the past few years; I regret that the timing of my ultimate defense conflicted with his schedule. I am indebted to Inés and Yin-Long for coming to my support in the eleventh hour (okay, 27th hour, but who is counting?). I wish I had taken advantage of the outside perspectives and modelling expertise of Aaron and Inés much earlier; the work presented here would have been the better for it.

I was able to discuss the ideas contained herein with numerous thoughtful people: Jonathan Bollback, David Cutler, Alexei Drummond, Doug Futuyma, Alan Gelfand, Oliver Haddrath, John Harshman, Simon Ho, Jeff Johnson, Paul Joyce, Aaron King, Karl Kjer, Vladimir Minin, Tom Near, Yin-Long Qiu, Michael Sanderson, Elliott Sober, Alexandros Stamatakis, David Swofford, Jeff Thorne, Marcel van Tuinen, Ziheng Yang, and Derrick Zwickl. These individuals served in several capacities, providing corrections, advice, information, data, encouragement, source code, or simply acting as a sounding board so that I could think out loud. I would like to acknowledge the support of my collaborators over the past 5 years: Ed Braun, Jaime García-Moreno, Briar Howes, Rebecca Kimball, Jeff Johnson, David Mindell, Robert Payne, Michael Defoin Platel, Yin-Long Qiu, Josh Rest, Michael Sorenson, Ulf Sorhannus, and Marcel van Tuinen. I thank Shing Hei Zhan for correcting my Latin grammar in the title of this dissertation. Thanks to Clay Cressler for assistance in plotting in R.

I am particularly indebted to two people. Paul Lewis, in addition to inspiring me to pursue molecular phylogenetics in the first place lo those many moons ago, provided thoughtful sentiments regarding statistical model selection criteria. Not only did Paul first introduce me to the model selection approach of Alan Gelfand and Sujit Ghosh (1998) (which unfortunately did not find its way in the present dissertation), but he also provided

some unpublished python source code that I was able to rewrite in C++ for my own program. No less deserving of praise, Bob Payne considerately read everything I have ever written, and in each instance offered sound advice for improvement (in particular, turning a shrill condemnation into a publishable scholarly 'comment'). Bob also furnished me with my first pair of binoculars, which in many ways altered my life trajectory.

Programming and computational phylogenetics are lonely pursuits. I thank the following for their steadfast companionship over the course of the past 5 years: Ron Asheton, Scott Asheton, Dave Alexander, David Bowie, Nick Drake, Robert Fripp, Kim Gordon, Polly Jane Harvey, Georgia Hubley, Ira Kaplan, György Ligeti, James McNew, Tim Minchin, Thurston Moore, Robert Pollard, Iggy Pop, Sun Ra, Lee Ranaldo, Lou Reed, Trent Reznor, Arnold Schoenberg, Steve Shelley, Dean Ween, Gene Ween, Frank Zappa, and especially, above everyone else, Mr. Don Van Vliet.

I would surely have gotten deported, audited, and/or thrown out of the program if it were not for the remarkable staff of the University of Michigan Museum of Zoology (UMMZ) and the Department of Ecology & Evolutionary Biology: Christy Byks-Jazayeri, Norah Daugherty, Beverley Dole, Janet Griffin-Bell, Janet Hinshaw, Gail Kuhnlein, Vlad Miskevich, Robbin Murrell, and Fritz Paper. I get easily confused by even simple paperwork; thank you all for your patience and for walking me through things. I am eternally beholden to the EEB graduate student coordinators that have helped my over the years: LaDonna Walker, Julia Eussen, and Jane Sullivan; Goddesses, all. Whatever salary they make, it is *way* too little.

The content of this dissertation was conceived, developed, analyzed, and written using Apple® computers. Part of this work (specifically, replicate MrBayes analyses for chapters 4 and 5) was carried out by using 1) the resources of the Computational Biology Service Unit from Cornell University (partially funded by Microsoft Corporation), 2) the Genomic Diversity Laboratory at UMMZ, and 3) the Center for Comparative Genomics PhyloCluster at the California Academy of Sciences. I thank Brian Simison at the California Academy of Sciences for granting me the privilege of acting as a beta tester on the PhyloCluster. Programming and scripting was composed using BBEdit (Bare Bones Software®) and compiled using the GNU Compiler Collection (gcc; Free Software

Foundation©). Figures not involving trees were plotted in R and/or Adobe Illustrator. Twinings® kept my cranium fueled and saturated with the delicious electrolytic solution of bergamot and black tea. How many cups of tea does it take to earn a PhD? 11,750.

All funding for this work was provided internally through various sources within the University of Michigan: University of Michigan Museum of Zoology, Department of Ecology & Evolutionary Biology, Rackham Graduate School, and the International Center. I am particularly thankful to Rackham for providing the very generous Predoctoral Fellowship which has provided all of my funding for the past year.

Graduate students with families do not have much time for extracurriculars. I don't think I could have better spent what little time I had available than with my friend George Kulesza. A PhD in ornithology I learned, at least. Surely there does not exist a more devoted, patient teacher. I count it amongst my highest achievements that you regard me as one of your 'successful' students.

Finally, I must acknowledge my family, the ultimate in extracurricular activities. Calliope and Nova are undoubtedly the coolest kids alive, and provide unlimited amusement. And lastly, my wife, Rachel. You have been around since this roller-coaster of graduate school began, and have been my supporter through the (all too infrequent) highs and the (all too frequent) lows.

Table of Contents

List of Tables

List of Figures

Part I. Temporal Calibration of the Avian Tree of Life

Chapter 1

Extraction of phylochronological information from molecular sequence data: evolution of our understanding of the 'evolution of the rate of evolution'[1]

INTRODUCTION

Evolutionary biologists are ultimately interested in biological diversity – how it is generated, how it is maintained, and how it is lost. Clearly our understanding of biodiversity would be incomplete without a temporal perspective. How long, for example, does it take for a clade to reach size *x*? When did novel adaptive innovation *y* evolve? What geophysical/environmental phenomena likely triggered speciation event *z*? It is here that the 'molecular clock' proves itself an extremely useful concept, as it allows elucidation of not only the timing of macroevolutionary events, but also the extent of their temporal clustering – what we might dub the 'phylochronological' signal. Extraction of this signal enables us to construct more informed hypotheses regarding the processes and mechanisms of diversification. Indeed, for taxa with poor or absent fossil records, a molecular clock may be the only means by which to infer phylochronological patterns. However, despite its utility, the 'molecular clock' concept has required considerable retooling to accommodate the heterogeneity ubiquitous to large molecular data sets (Bromham and Penny, 2003; Magallón, 2004; Rutschmann, 2006; Welch and Bromham, 2005). On the whole we can perceive a trend to making molecular clock models more general by relaxing simplifying assumptions of previous implementations.

---

[1] Published as Brown, J. W., and M. van Tuinen. 2010. Evolving Perceptions on the Antiquity of the Modern Avian Tree *in* The Evolutionary History of Modern Birds (G. J. Dyke, and G. Kaiser, eds.). UC Press. All material presented here written by JWB.

ORIGINS OF THE MOLECULAR CLOCK

Molecular clock theory was borne of the pioneering work of Emile Zuckerkandl and Linus Pauling (Zuckerkandl and Pauling, 1962; 1965), and Emanuel Margoliash (1963). Working with mammalian protein sequences, these authors remarked that sister lineages contained very similar numbers of amino acid substitutions. From these empirical observations they posited that although molecular substitution is best regarded a stochastic process, over long periods of time it can be considered approximately constant at rate λ. This simple yet powerful hypothesis yields testable predictions, which, if passed, enables us to interpret a molecular phylogeny in terms of absolute time rather than a simple nesting of clades. Assuming a strict molecular clock, dating of nodes in a tree is a trivial task. The time to the most recent common ancestor, t_{MRCA}, of two taxa separated by genetic distance d is calculated as:

$$t_{MRCA} = \frac{d}{2\lambda}$$

(the coefficient 2 is required because both lineages undergo substitutional accumulation in time t). This equation assumes, of course, that the (constant) rate of molecular evolution λ is known. The value of λ usually comes from the calibration of genetic distances with the fossil record (see below), although occasionally dates of biogeographic events are used in place of fossil calibrations. Extending this to multiple taxa is straightforward, since all lineages within the tree are assumed to share the same value of λ, although there will generally be a need to correct for stochastic deviations from ultrametericity (i.e. that all terminal branches in the phylogram line up at the present). Unfortunately, most present day molecular data sets reject the economy of the strict molecular clock. Avian dating studies wishing to employ a global molecular clock therefore require data pruning, either via 'gene-shopping' (Hedges et al., 1996; Kumar and Hedges, 1998) or 'taxon-shopping' (van Tuinen and Dyke, 2004; van Tuinen and Hedges, 2001), to obtain a matrix that will not reject a molecular clock. This is a reasonable practice if one believes that the majority of genes and taxa conform to expectation. However, this is unreasonable if one wishes to retain all hard-earned data within an analysis, or if one believes that the processes of substitution are more

heterogeneous. Indeed, on the timescale of neornithine evolution one might predict that the signal of a strict molecular clock would decay due to stochastic variations alone. I will briefly summarize here the most popular approaches to estimating divergence times with non-clocklike data (Table 1.1).

ACCOMMODATING NON-CLOCKLIKE MOLECULAR GENETIC DATA

1. Overdispersed Clocks

Rejection of a molecular clock is typically considered as evidence for rate variation across lineages. However, as Gillespie and Langley (1979) argue, the molecular clock hypothesis (as commonly employed) actually consists of two constituent assumptions: 1) that substitution rates are constant, and 2) that substitutions occur according to a Poisson process. An alternative interpretation of lineage-specific variability in the number of substitutions could thus be to question the second assumption; specifically, whether the variance afforded through a Poisson distribution (where the variance is equal to the mean) of substitutions through time adequately describes the variability we observe in empirical data sets. This interpretation is validated through use of alternative Gaussian (Cutler, 2000) and negative binomial (Bedford et al., 2008) distributions which adequately describe lineage-specific substitution counts where a Poisson distribution fails. An interesting corollary of this alternative constant-rate high-variance molecular clock hypothesis is the absence of any assumed correlation of rate with phylogeny (see below); in an overdispersed clock long branches need not be clustered on a tree, because 'rate of evolution' is not assumed to be a heritable trait (a strict molecular clock, in contrast, carries with it the implicit assumption that substitution rate is entirely heritable). Indeed, application of Cutler's (2000) method (as implemented in the program dating5) to avian mtDNA sequence data revealed distinct temporal diversification patterns not revealed through other methods (Brown et al., 2008). Whereas autocorrelated methods (see below) generally infer gradual patterns of diversification, the overdispersed reconstruction infers both short periods of extensive diversification and long periods of stasis. Despite the perceived promise and interpretive simplicity of an overdispersed clock, existing analytical programs are limited and thus are rarely used.

2. Local Clocks

The more common approach to the rejection of a Poisson-distributed molecular clock is to assume among-lineage rate heterogeneity. The most straightforward way to extend the original molecular clock concept to accommodate rate variation across a tree is to employ 'local' molecular clocks (Yoder and Yang, 2000). Here, regions of a tree are assumed to evolve according to a strict Poisson-distributed molecular clock, but different clades can have different rates λ_i. This is certainly a better description of empirical data, as different clades of birds display markedly different rates of molecular evolution as revealed through trends in branch length heterogeneity in reconstructed phylograms (e.g. Hackett et al., 2008). However, the number of individual local clocks and their discrete placement in the tree is inherently subjective. Although various local clock models can be compared statistically (in the program PAML; Yang, 2007), the local clock approach has largely been abandoned for approaches that let the data themselves indicate where changes in evolutionary rate likely occur within a tree.

3. Rate Smoothing

The local clock approach above assumes a few (potentially great) discrete changes in the Poisson rate of substitution λ_i across a tree. A more general approach is to allow an arbitrary number of such changes in λ_i, but to 'smooth' transitions in rate to minimize large changes. Two general approaches exist, these differing in the direction of smoothing; it is debatable which direction is optimal. On the one hand, sister lineages are by definition the exact same age, and because they share a recent common ancestor they are likely of similar size, life history traits, DNA repair efficiency, etc. – characteristics that are thought to influence substitution rates. It therefore seems sensible to focus on sister lineages when minimizing deviations from a molecular clock. This is the approach that PATHd8 (Britton et al., 2007) takes, and can be regarded as a smoothed local-clock approach. Here, path lengths are averaged successively from the tips of a tree back through internal nodes. The averaged sister path lengths are assumed to obey a strict Poisson molecular clock, although different sister-pairs can have different rates because of branch length differences or reference to simple fossil-imposed age constraints. The simplicity of the calculations involved allow for the dating of very large trees (hundreds

of taxa), very quickly (typically << 1 second of computation). However, when applied to avian data (Ericson et al., 2006), the approach infers divergence time estimates that are strikingly younger than those from alternative more rigorous approaches (Brown et al., 2007; 2008), suggesting that the method may be overly simplistic. Indeed, the current implementation of PATHd8 has been demonstrated to be statistically biased, generating overly young and precise divergence time estimates for even relatively simple simulated sequences (Svennblad, 2008). More work is required to see if this method can be rescued, but at the moment it is best considered with caution.

The alternate direction of smoothing, from ancestor to descendant branches, may therefore be more reasonable; indeed, the evolution of rate variation from ancestor to descendant branches is the very process we are trying to understand. If the trait 'rate of molecular substitution' is heritable to any degree (for instance, because of inheritance of DNA-repairing enzymes), then smoothing in this direction can be expected to extract more meaningful information. The program r8s (Sanderson, 2003) takes such an autocorrelated-rates approach, but penalizes rates that change too quickly across the tree in a fashion akin to smoothing in regression analysis. In non-parametric rate smoothing (NPRS; Sanderson, 1997) optimal rates and dates are inferred by simply minimizing the penalty function. However, NPRS is generally not recommended for most data sets as it tends towards overfitting, inferring large fluctuations in rate where short branch lengths are located. The alternative semi-parametric penalized likelihood (PL; Sanderson, 2002) approach is an extension to NPRS which involves a smoothing parameter which controls the relative contributions of rate smoothing and data-fitting; large values of the smoothing parameter favour minimizing rate changes over data-fitting (tending towards a molecular clock), while small values of the smoothing parameter tend towards NPRS. The optimal smoothing value is determined through a data-driven sequence-based cross-validation procedure. Application of this method to avian data matrices has yielded reasonably consistent divergence time estimates that generally agree with more realistic, computationally intensive approaches (Baker et al., 2005; Brown et al., 2008; Haddrath and Baker, 2001; Harrison et al., 2004; Paton et al., 2002). Nevertheless, the method (as currently implemented) has the drawback that a single global smoothing parameter controls the extent of rate change across an entire tree. This assumption may be

unreasonable for trees that potentially span tens or hundreds of millions of years if constraints on substitution rate variability have changed appreciably over evolutionary time. Likewise, the autocorrelation assumption that forms the basis of r8s smoothing has recently come under question (Drummond et al., 2006), and for birds specifically (Brown et al., 2008), although the extent of the decay of autocorrelation of rates is likely dependent upon taxon sampling and tree age.

4. Modelling 'Relaxed' Clocks

A more rigorous approach to dating non-clocklike molecular genetic sequence data is to explicitly model rate heterogeneity itself. Although much more computationally time-consuming than the methods above, these model-based approaches have two basic advantages: 1) they extract more biologically-interpretable information, and 2) the relative fit of alternative candidate relaxed-clock models to empirical sequence data can be computed statistically using existing tools. Modelling rate heterogeneity generally takes one of two forms: 1) modelling the *process* of rate change across a tree, or 2) modelling the *product* of such rate changes. The former methods require assumptions about how rate change proceeds over evolutionary time. The most highly utilized model of this type was developed through the work by Jeff Thorne and colleagues on modelling the "the rate of evolution of the rate of evolution" (Kishino et al., 2001; Thorne and Kishino, 2002; Thorne et al., 1998). Implemented in the popular MCMC program Multidivtime (Thorne, 2003), this model implicitly assumes an autocorrelated process of rate evolution from ancestor to descendent branches. Specifically, the model assumes that the substitution rate at the descendent branch conforms to a lognormal distribution, the mean of which is equal to the logarithm of the rate at its ancestral branch. The variance of this lognormal distribution is determined by both a sampled autocorrelation parameter ν and the inferred length of time separating the two nodes. An intuitively satisfying property of this approach is that sister branches, while being autocorrelated (to some degree) in rate to their common ancestral branch, can nevertheless potentially differ considerably from one another. This model has been applied extensively to avian data matrices (Baker et al., 2007; Brown et al., 2007; Brown et al., 2008; Pereira and Baker, 2006a; Pereira and Baker, 2006b; Pereira and Baker, 2008; Pereira et al., 2007; Slack et

al., 2006), and is largely responsible for the statistically robust Cretaceous molecular timescale that has been emerging over the past decade (Figure 1.1). Nevertheless, Multidivtime is starting to show its age, being limited to a simple substitution model (F84+G) which is inappropriate for large taxon and/or character samples, and simple 'hard' age calibrations (upper, minimum, or fixed; see below).

Alternative approaches to modelling autocorrelated-rates across a tree are based upon the Ornstein-Uhlenbeck (OU) process. Also called a 'mean-reverting process', here rates can change across a tree in a near-Brownian fashion, although an equilibrium rate is enforced through use of a 'spring' restraint that 'pulls' a rate (either down or up) towards the equilibrium rate with a force proportional to how far it is removed from the mean; in effect, the model penalizes extreme rates. Unlike the autocorrelated lognormal model of Multidivtime, the OU model (and its variants; see below) possesses a stationary distribution (i.e. the mean and variance do not change over time or across the tree), and hence is a very different conceptual take on the process of rate evolution; whereas in Multidivtime a branch-specific rate is explicitly tied to its ancestral branch rate, in methods employing the OU process all rates are instead tied to the same underlying equilibrium rate. The idea of the reality of an 'underlying equilibrium rate' is a deep and provocative assumption about the process of rate evolution, and in a way the OU approach can be thought of as modelling the distribution of rates around an 'absolute' molecular clock. An early implementation of the OU process for phylochronological analysis (in the program PhyBayes; Aris-Brosou and Yang, 2002; Aris-Brosou and Yang, 2003) was shown to be flawed through inappropriate priors overly influencing the results (Welch et al., 2005); in particular, the priors were biased to infer higher rates of substitution near the root of the tree. However, the recent implementation of the 'CIR' model (essentially a 'squared-OU' model, which preserves rate positivity and avoids the prior bias above) in the program PhyloBayes overcomes many of these problems, and is well supported by empirical data (Lepage et al., 2007). Despite the promise and success of this approach, its very recent development has meant that avian molecular genetic data has yet to be analyzed in this way.

While the models above represent autocorrelated rate evolution as a continuous process, it is also possible to build piecewise relaxed clock models. For example, the compound Poisson stochastic process approach of Huelsenbeck et al. (2000) assumes that substitutions generally occur according to the standard Poisson-distributed molecular clock, but that changes in rate λ_i occur along the tree according to an independent Poisson process. Using standard MCMC machinery, divergence times and their associated credible intervals can be estimated while accommodating uncertainty in all other model parameters, including the frequency and degree of discrete rate changes. This model can be thought of as a generalized local clock approach, but has an advantage over the subjective procedure above in that the number, degree and location of the inferred shifts in substitution rate are data-driven instead of investigator-proclaimed. Unfortunately, a lack of development beyond its introductory paper (Huelsenbeck et al., 2000), together with a lack of available software (but see Himmelmann and Metzler, 2009), has meant that this straightforward and biologically-interpretable approach has yet to realize its potential.

In contrast to the models above, the second class of models, those that model the *product* of rate heterogeneity, do not make any explicit assumptions about how rate changes. Rather, these models make assumptions about the shape of the resulting distribution of rates, and assume that branch-specific rates are each drawn independently from this distribution. Using MCMC methodologies, proposed branch rates are accepted at a frequency that is proportional to their posterior probabilities. Unlike the other models above (but similar to the overdispersed clock), these models make no assumptions regarding an autocorrelation of substitution rates across a tree, and so are frequently referred to as 'uncorrelated' models. There is good reason to question the autocorrelation assumption; even if 'rate of evolution' *is* heritable, the accumulation of stochastic variation over millions of years may mean that autocorrelation decays to zero along the branches separating the nodes in a tree (Drummond et al., 2006). Regardless, relaxing the autocorrelation assumption means that autocorrelation itself can be tested; if rates are indeed autocorrelated (and sufficient signal is present in the data) then the sampled rates should reflect that pattern. Autocorrelation has only been evaluated once in a broad scale sample of Neornithes (Brown et al., 2008), and was rejected. However, because of the

time-dependency of autocorrelation decay, recovery of a genuine signal of autocorrelation will likely require a more dense taxon sampling than has been performed previously so that time intervals between nodes can be minimized.

Among these uncorrelated models, the uncorrelated lognormal model has enjoyed the most use to date, and has been shown to be superior to an uncorrelated exponential model of rate variation for a number of data sets (Drummond et al., 2006), including birds (Brown et al., 2008). The flexibility of the lognormal distribution means that it is able to be fit to a broad range of rate distribution shapes, and explains why it is implemented in several Bayesian relaxed clock applications, including BEAST (Drummond et al., 2006; Drummond and Rambaut, 2007), MCMCtree (Rannala and Yang, 2007; Yang, 2007), and PhyloBayes (Lartillot and Philippe, 2004; Lepage et al., 2007). Each of these packages has its own advantages. MCMCtree, for example, explicitly allows for potential error in fossil calibration ages (for example, from stratum misidentification) by adding non-zero probability tails to otherwise 'hard' fossil constraints (Yang and Rannala, 2006). The benefit of using PhyloBayes is that it implements seven different clock models, enabling a researcher to statistically compare alternative models using the same statistical machinery rather than relying on indirect comparisons across software packages/implementations (Lepage et al., 2007). Finally, in addition to flexible xml-coding support which allows for the construction of arbitrarily complex models, BEAST is unique in that of all the relaxed clock methods available, it is the only one that does not require a fixed tree topology. This inclusion of topological uncertainty is especially appealing for avian studies, where higher level relationships are still unsettled.

In summary, there are currently a number of approaches readily available to researchers for phylochronological reconstruction using non-clocklike molecular genetic sequences (Table 1), although none of them can be considered a panacea (see below). These approaches run the gamut from quick-and-dirty 'corrections' to an imperfect clock (e.g. PATHd8) to sophisticated descriptions of either the process of rate evolution itself (e.g. CIR model) or the product of such evolution (e.g. uncorrelated lognormal model). Given the breadth of choices available, the ideal course of action would be to test several distinct approaches to see if phylochronological signal is consistent across model/method

assumptions (Britton et al., 2007; Brown et al., 2008; Hipsley et al., 2009; Hug and Roger, 2007; Lepage et al., 2007; Linder et al., 2005): concordant results across methods would lend additional credence to resulting inferences, whereas dissonance could help identify potential model assumption violations. For example, a recent comparison of five dating methods on the Neornithes tree using mtDNA revealed broadly consistent origin estimates for the major clades (Brown et al., 2008), while also calling into question the appropriateness of one method (PATHd8) for the particular data set.

LIMITATIONS OF CURRENT MOLECULAR PHYLOGENETIC DATING TECHNIQUES

A model represents a conceptual understanding of how 'nature' influences physical entities to generate the distribution of empirical observations. Models can be constructed from empirical (inductive) or theoretical (deductive) expectations, with the ideal situation being a motivated iteration between the two sources of understanding (Box, 1976). However, a model should not endeavor to 'fit an elephant' (that is, try to describe reality in its entirety; Steel, 2005), but instead attempt to extract information from the *salient* components of the underlying process, formalized with estimable parameters. The idea of *saliency* should be recognized as a relative concept; with greater thought, and a broader collection of empirical observations, our idea of what constitutes a 'salient' component of a process continues to evolve, leading to a richer understanding of the sources of variation. Such is the condition of our understanding of the processes of molecular evolution. Larger molecular genetic data matrices (in terms of both taxon and especially character sampling) has afforded an increased power to identify more subtle (but increasingly important) sources of variation. Consequently, several simplifying assumptions in our standard modelling of the molecular genetic evolutionary process are currently being challenged, and may eventually translate to improvements in the extraction of phylochronological signal or potentially identify biases of past methods.

1. Molecular Substitution Models

Molecular evolution is typically modelled as a continuous-time Markovian substitution process. This conceptual framework, originally constrained for practical (i.e. computation tractability) reasons (Felsenstein, 1981), carries with it several explicit and implicit

assumptions: 1) *stationarity* (the probabilities of stochastic substitution do not change through time, or are in equilibrium); 2) *homogeneity* (the equilibrium character frequencies and substitution-rate matrix are identical across lineages); 3) *time-reversibility* (the process of substitution looks the same both forwards and backwards in time); and 4) *independence* (all sites within an alignment are considered identical and independently distributed (i.i.d.) realizations of the same evolutionary process). Although held by all of the relaxed clock methods above, none of these assumptions is likely to strictly hold true (and indeed empirical data exist to contest each of them), however since all models are wrong we should concern ourselves with what is *importantly* wrong (Box, 1976).

Of these assumptions, *independence* is unique in that it focuses along a genetic sequence (rather than across a tree like the remaining assumptions). This assumption is actually a composite assumption: 1) sites evolve independently from one another, and 2) all sites evolve according to the same underlying process (in practice, the same substitution model). Strict violation of the first component is ensured through the physical linkages between nucleotides, although inclusion of molecular markers from disparate regions of the genome (say, different chromosomes) can represent 'more independent' information. Violation of the second component is readily apparent through inspection of the characteristics of various character classes (e.g. genes, coding/non-coding regions, codon positions, etc.) which often differ considerably in terms of nucleotide composition and levels of polymorphism. Failure to accommodate for this will necessarily lead to a 'compromised' inference (where, for example, relative rate parameters and equilibrium character state frequencies are averages over potentially distinct genomic regions). Nevertheless, the *independence* assumption is also unique in that its violation is all but solved. For example, the introduction of among-site gamma-distributed rate heterogeneity enormously increases the fit of models to empirical data (Yang, 1996). More generally, recently developed mixed (Lartillot and Philippe, 2004; Pagel and Meade, 2004) and partitioned (e.g. Nylander et al., 2004) models allow heterogeneity in the substitutional process across sites and loci. Partitioned models are available in a number of relaxed clock methods, although they are most flexible in BEAST.

The remaining model assumptions above reflect expectations of the uniformity of the molecular substitution process(es) over both time and lineages. The adoption of the *time-reversibility* assumption is due almost entirely to numerical convenience: it both reduces the required number of substitution parameters to be estimated (as compared to the more general model; Rodríguez et al., 1990), and allows for efficient computation of the likelihood of an unrooted tree via Felsenstein's 'pulley principle' (Felsenstein, 1981). However, reasons to settle for simpler reversible models are rapidly dematerializing: 1) Bayesian MCMC-sampling methods, together with increasingly powerful and available computational resources, can easily accommodate the relatively small increase in the number of estimable parameters, and 2) signal from large present day empirical data matrices are revealing support for irreversibility (Squartini and Arndt, 2008), overturning earlier conclusion from studies (Yang, 1994) that may simply have suffered from a lack of power.

The final two assumptions, *homogeneity* and *stationarity*, are tightly related in that violation of one typically involves violation of the other. Of all the assumptions, these two are most likely to influence phylochronological inference through biasing both topology and branch length estimation. Moreover, the strict validity of these assumptions over evolutionary time spans, such as the diversification history of Neornithes, is dubious. Indeed, clear evidence for the violation of one or both of these assumptions is the observed empirical base compositional biases across lineages that can not be described by stochastic variation alone. While such compositional biases can potentially be masked through data filtering (e.g. translating to amino acids for coding sequences, or employing R-Y coding), a more satisfying approach that makes use of more evolutionary information is to model compositional changes themselves. Fortunately, several non-homogeneous/non-stationary models exist which do just that (e.g. Galtier and Gouy, 1998), for example by allowing base composition to change across a tree according to a compound piecewise-constant Poisson stochastic process (Blanquart and Lartillot, 2006), similar to the compound Poisson relaxed molecular clock model (Huelsenbeck et al., 2000) above. A distinct violation of the stationarity assumption involves the concept of 'heterotachy', where site-specific substitution rates change in different parts of the tree (Lopez et al., 2002). Thankfully, substitution models now exist that accommodate

heterotachous evolution (Kolaczkowski and Thornton, 2008; Pagel and Meade, 2008; Tuffley and Steel, 1998; Wu et al., 2008). In summary, it is not yet known whether biologically realistic violations of these substitution model assumptions will surface as significant biases to phylochronological signal. However, we have at our disposal solutions to each of these problems, all that remains is to graft constituent models together (e.g. Blanquart and Lartillot, 2008), identify the salient features, and apply the resulting models to the problem of phylochronological inference.

2. Modelling of Rate Evolution

Besides concerns regarding the suitability of the level of sophistication of existing molecular substitution models, we must also consider whether the correct components of evolutionary rate heterogeneity are indeed being modelled. It is unclear, for example, if substitution rate evolution is best considered autocorrelated in time (such that 'rate' is a heritable character), or if an 'episodic' clock (Gillespie, 1984) is a better description of empirical data. Correlated rate models, if valid, enable greater inferential precision because rate/date estimation at a given node can make use of not only local but also distant evolutionary information (Lepage et al., 2007). The uncorrelated models above are episodic clocks that offer no explanation of *why* rates vary. A distinct type of episodic clock involves punctuated (or speciational) molecular evolution, where substitution rates are elevated during speciation, with the result that the lengths of the branches (in terms of the expected number of substitutions per site) in a clade of a tree are positively correlated to the number of speciation events (Pagel et al., 2006). Such a scenario could explain stark branch length differences between the speciose Passeriformes and depauperate Pelecaniformes in reconstructed phylograms (Hackett et al., 2008). Indeed, punctuated *morphological* evolution has been inferred in birds (Paleognaths; Cubo, 2003). However, a signal of punctuated *molecular* substitution rates (which could potentially mislead phylogenetic dating) was not found in a recent study of Neornithes (Brown et al., 2008), although identification of this kind of signal would surely benefit greatly by increased taxon sampling. Another potential explanatory variable to consider is effective population size, which has been shown to be correlated (negatively) with substitution rate in a range of eukaryotic taxa (Bedford et al., 2008). Additionally, effective population size strongly

determines the rate of lineage sorting, and since lineage sorting postdates speciation it makes sense to estimate divergence times and effective population sizes simultaneously (Liu and Pearl, 2007; Rannala and Yang, 2003). Given that many avian species differ greatly in effective population size (think of rails versus gulls), this may be a worthwhile avenue of research to pursue. A final consideration involves the perceived time-dependency of molecular substitution rates (Ho and Larson, 2006; Ho et al., 2005); here, extant population-level polymorphisms that would not persist over evolutionary timescales (i.e. mutations that do not become substitutions) bias reconstruction methods into inferring that substitution rates are higher in the present than they were in the past. The influence of this pervasive phenomenon on divergence time estimation has not yet been fully investigated, and not at all in birds.

3. Age Constraints

One of the most compelling developments in recent relaxed clock model implementations is the ability to construct age probability distributions for fossil-calibrated nodes. Previous molecular dating techniques (e.g. r8s, Multidivtime, and PATHd8) allowed only the enforcement of 'hard' age constraints: 1) absolute minimum (i.e. that the speciation event represented by the calibrated node *must* predate the fossil; the fossil, of course, being a product of the speciation event), 2) absolute maximum (information which, strictly speaking, can not come from the fossil record), or 3) fixed ages (i.e. the fossil perfectly represents the age of the node without error). The new probability distributions available in BEAST (Drummond and Rambaut, 2007) and MCMCtree (Yang and Rannala, 2006) offer two main advantages over these simple constraints: 1) additional information (e.g. from models of fossil preservation) can be incorporated into the calibration, effectively lending more credence to the fossil record, and 2) uncertainty in the age of the fossil itself can be accommodated. However, the same flexibility that makes these distributions so attractive unfortunately also makes them inherently subjective. Although rightly considered with enthusiasm (Ho, 2007; Ho and Phillips, 2009), there is presently no rigorous protocol for determining the optimal shape (e.g. Gaussian, lognormal, uniform, exponential, etc.; Figure 1.2) and breadth of these distributions, which makes direct comparisons across studies difficult. A joint

collaboration of paleontologists and molecular phylogeneticists working on this problem would allow greater extraction of phylochronological signal from the fossil record, and subsequently generate better divergence time estimates.

4. Study Design

Finally, improvements in divergence time estimation will require systematic attention to sampling with respect to which loci, taxa, and fossil calibrations should be included in a given study. For example, over long evolutionary timescales mtDNA can be expected to exhibit substitutional saturation, which may bias relaxed clock studies through an underestimation of branch lengths deep in the tree (consequently underestimating the ages of deeper nodes). Under this scenario, it may be desirable to utilize slower evolving nuclear introns, or to mask saturation through translating nucleotide sequences to amino acids (if coding) or otherwise through RY-recoding (Woese et al., 1991). The extent of taxon sampling has been found to be influential in molecular dating (e.g. Linder et al., 2005), presumably due to node-density effects (Venditti et al., 2006), where more substitutions are discovered (making branch lengths longer) in regions of the tree with higher taxon sampling. Lastly, while it is generally a good strategy to incorporate calibration information from as many fossils as possible (Bremer et al., 2004; Hug and Roger, 2007), it is imperative that these fossils are scrutinized closely, as one incorrectly-dated or taxonomically misdiagnosed fossil can potentially invalidate an entire analysis. It thus seems prudent to test suites of calibrations for dating consistency (Near et al., 2005; Near and Sanderson, 2004), although it should be kept in mind that exceptionally 'good' fossils (i.e. those that are especially old, or more closely approximate the age of the node they are meant to date) are likely to appear 'inconsistent'.

In conclusion, we are at a very exciting stage in molecular phylogenetic systematics; not only are we well aware of the potential unsuitability of assumptions made by early relaxed clock approaches, but (more importantly) we have a firm grasp on what *further* tribulations may be lurking in the future. The widespread adoption of Bayesian philosophies over the past decade in particular has ushered in a new paradigm for methodological implementation, *complecto errorem* (embrace uncertainty), where uncertainty in 'nuisance' parameters (model components that are essential for a salient

description of the evolutionary process, but are otherwise not of direct interest) can be integrated out, rendering conclusions that are compelling to a degree that have not been heretofore possible. Moreover, the rapidly-decreasing costs associated with molecular genetic sequencing means that it will soon be possible to interrogate enormous amounts of data for subtle signals of past molecular substitution rate evolution. With such information in hand we can expect more accurate, precise, and consistent phylochronological inferences, which in turn will better enable us to understand and appreciate the dynamics of neornithine diversification.

ACKNOWLEDGEMENTS

I thank A. Drummond for helpful discussions, J. Johnson and J. McCormack for comments on earlier drafts of the manuscript, and G. Ligeti and S. Youth for encouragement throughout.

Table1.1 A comparison of available programs for the dating of non-clocklike molecular genetic sequences.

Program	Clock approach	Statistical inference	Input data	Multiple partitions	Age constraints	Topology	Notes
PATHd8	rate smoothing (sister branches)	N/A	phylogram	no	minimum, maximum, and fixed[1]	fixed	extremely fast for even large trees; current implementation delivers overly young/precise estimates
r8s	rate smoothing (ancestor-descendent)	penalized likelihood	phylogram	no	minimum, maximum, and fixed	fixed	optimal smoothing via sequence-based cross-validation
dating5	overdispersed clock	maximum likelihood	phylogram	no	minimum, maximum, and fixed	fixed	software has not seen recent development
PAML	local clocks	maximum likelihood	nucleotide or amino acid sequences	yes	minimum, maximum, and fixed	fixed	location of discrete clocks is user-defined
HLS2000[2]	compound Poisson process	Bayesian MCMC	DNA sequences	no	fixed[1]	fixed	software has not seen recent development
Multidivtime[3]	autocorrelated model	Bayesian MCMC	nucleotide or amino acid sequences	yes	minimum, maximum, and fixed	fixed	nucleotide substitution model limited to F84; too simplistic for most large/old trees
PhyBayes	OU process	Bayesian MCMC	nucleotide sequences	yes	N/A[5]	fixed	software has not seen recent development
MCMCtree[4]	uncorrelated lognormal	Bayesian MCMC	nucleotide or amino acid sequences	yes	probability distributions	fixed	explicitly accommodates potential error in calibration ages
PhyloBayes	various autocorrelated and uncorrelated models	Bayesian MCMC	nucleotide or amino acid sequences	no	N/A[5]	fixed	7 clock models can be compared within the same software package
BEAST	uncorrelated lognormal or exponential	Bayesian MCMC	nucleotide or amino acid sequences	yes	probability distributions	estimated	xml-coding allows for arbitrarily complex models of sequence evolution

[1] A fixed-age constraint is required.
[2] This program was not given a proper name, so the initials of the authors are used. Not much is known about this program as it is not distributed and has not enjoyed use beyond the original study. Although a number of possible extensions are discussed by the authors, these apparently have not been implemented.
[3] The initial steps of analysis require the PAML package.
[4] Part of the PAML package.
[5] Estimates relative ages. Absolute ages are generated through scaling relative ages to a fixed age constraint; if more than one constraint is available an average date across all constraints is estimated.

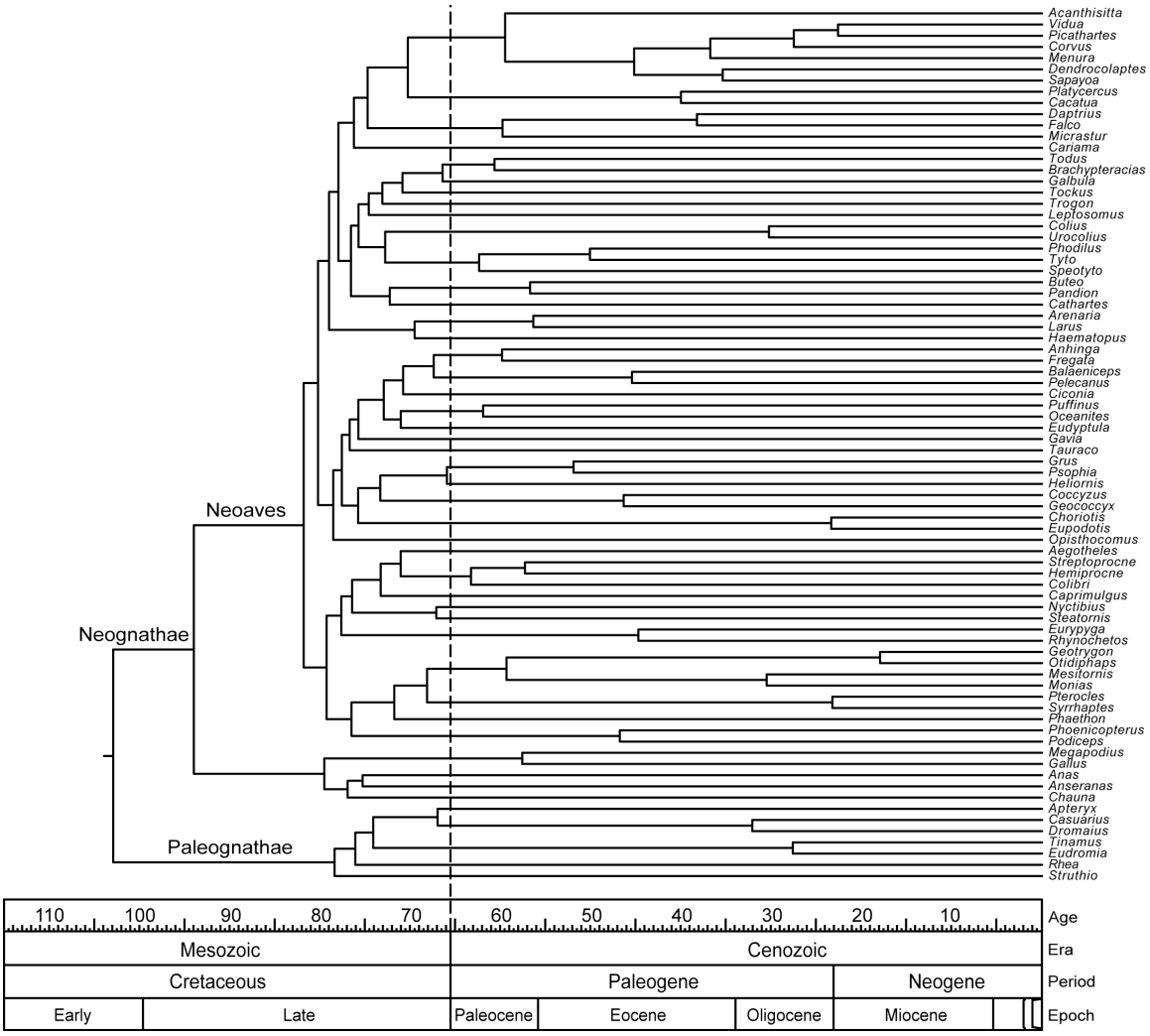

Figure 1.1 Molecular genetic evolutionary timescale of neornithine diversification inferred from the analysis of data from Hackett et al. (2008) utilizing a Bayesian uncorrelated lognormally-distributed relaxed molecular clock in BEAST (Brown et al., unpublished data). The taxon matrix was pruned to include only the major neornithean lineages. The vertical dashed line identifies the Cretaceous-Paleogene (K-Pg) boundary. Node ages are mean values from the posterior distribution; credible intervals are omitted for clarity.

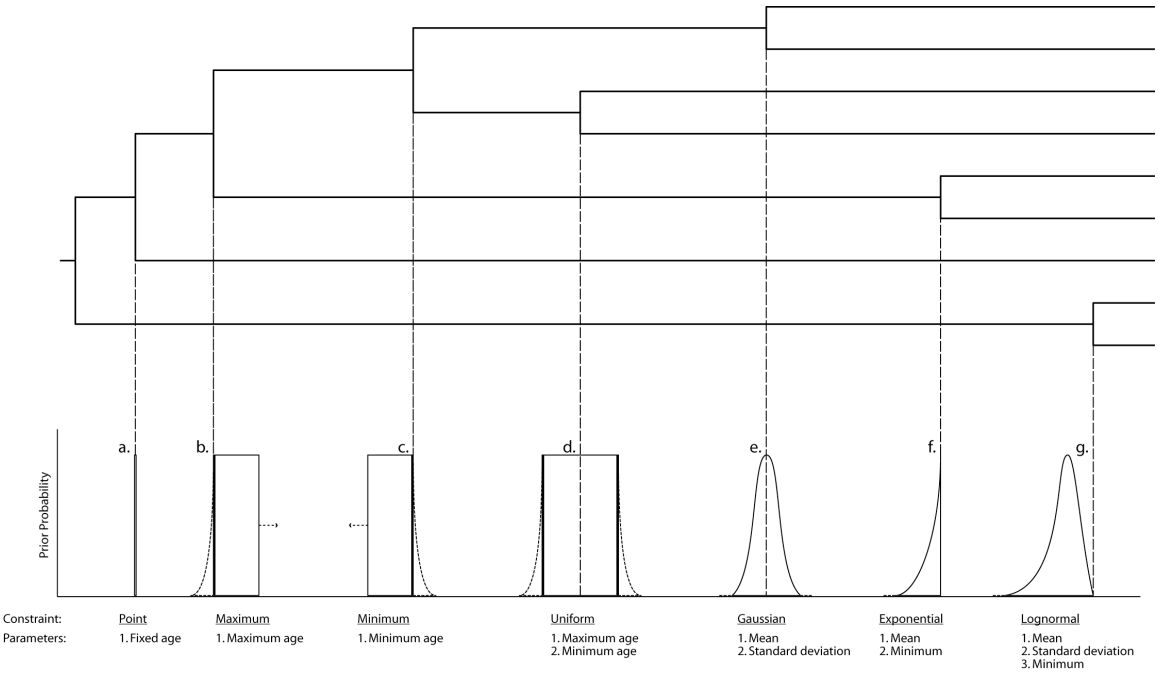

Figure 1.2 Alternative temporal calibration constraints for use in molecular divergence time estimation. (a) At the extreme, the age of a fossil may be used as a point estimate for the divergence of two taxa. Here, the age of the node is treated as known without error; hence, potential divergence times above or below the point estimate receive zero prior probability, and so are not considered in the estimation process. This is generally poor practice, as 1) potential errors in phylogenetic placement and geological dating of the fossil are ignored, and 2) diagnosable fossils are unlikely to temporally correspond precisely to cladogenesis. Alternatively, fossils can be considered as (b) maximum or (c) minimum ages. [Strictly speaking, fossil evidence can only provide minimum ages, not maximum ages, as the absence of (earlier) evidence cannot necessarily be interpreted as evidence of (earlier) absence]. Typically, such calibrations are implemented as 'hard' constraints (bold lines), where prior divergence times on one side of the constraint are equiprobable (arrows) while all values on the other side of the constraint receive zero prior probability. However, such calibrations can also be implemented as 'soft' constraints (Yang and Rannala, 2006), where exponential distributions constituting some proportion of the overall prior density (say, 2.5%) extend beyond the 'hard' bounds (dashed curves). A soft prior gives small but non-zero probability to ages beyond that of the fossil and hence accommodate potential errors in geologic dating or phylogenetic placement, potentially leading to the identification of inappropriate/invalid constraints. Minimum and maximum constraints can be combined into a single calibration. (d) When considerable uncertainty is present this is typically modelled with a uniform (equiprobable) distribution (with either 'hard' or 'soft' bounds). (e) In situations where existing evidence suggests some time periods are more probable *a priori* than others (say, biogeographic events), a more appropriate modelling may take the form of a Gaussian (normal) distribution. (f) The form of the exponential distribution is such that the mode (highest probability) of the distribution is the age of the fossil itself; ages younger than the fossil receive zero prior probability (i.e. a 'hard' bound). Application of an exponential constraint is useful when 1) a fossil is thought to temporally correspond closely to the relevant cladogenetic event, or 2) where additional information (say, fossil preservation curves) can be used to inform prior construction. (g) Perhaps the most appropriate distribution to model uncertainty regarding a cladogenetic event is the flexible lognormal distribution. This distribution has a 'hard' minimum at the age of the fossil, a lag until the mode of the distribution (i.e. sampled diagnosable fossils are expected to post-date the cladogenetic event), and has a long ('soft') tail that allows for the (small but non-zero) possibility that the divergence event occurred much earlier than what the fossil calibration would suggest. As with the Gaussian (standard deviation) and exponential (where mean = standard deviation) distributions, the breadth of the lognormal distribution is considerably malleable to the information at hand. Nevertheless, at the moment construction of lognormal (and other) constraints is more of an art than a science; scientists would do best to investigate the sensitivity of divergence time inferences to changes in temporal prior constraints.

REFERENCES

Aris-Brosou, S., and Z. Yang. 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.

Aris-Brosou, S., and Z. Yang. 2003. Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. Molecular Biology and Evolution 20:1947-1954.

Baker, A. J., L. J. Huynen, O. Haddrath, C. D. Millar, and D. M. Lambert. 2005. Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant moas of New Zealand. Proceedings of the National Academy of Sciences of the United States of America 102:8257-8262.

Baker, A. J., S. L. Pereira, O. P. Haddrath, and K.-A. Edge. 2006. Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. Proceedings of the Royal Society of London B Biological Sciences 273:11-17.

Baker, A. J., S. L. Pereira, and T. A. Paton. 2007. Phylogenetic relationships and divergence times of Charadriiformes genera: multigene evidence for the Cretaceous origin of at least 14 clades of shorebirds. Biology Letters 3:205-209.

Bedford, T., I. Wapinski, and D. L. Hartl. 2008. Overdispersion of the molecular clock varies between yeast, drosophila and mammals. Genetics 179:977-984.

Benton, M. J. 1999. Early origins of modern birds and mammals: molecules vs. morphology. BioEssays 21:1043-1051.

Blanquart, S., and N. Lartillot. 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. Molecular Biology and Evolution 23:2058-2071.

Blanquart, S., and N. Lartillot. 2008. A site- and time-heterogeneous model of amino acid replacement. Molecular Biology and Evolution 25:842-858.

Box, G. E. P. 1976. Science and statistics. Journal of the American Statistical Association 71:791-799.

Bremer, K., E. M. Friis, and B. Bremer. 2004. Molecular phylogenetic dating of asterid flowering plants shows early Cretaceous diversification. Systematic Biology 53:496-505.

Britton, T., C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer. 2007. Estimating divergence times in large phylogenetic trees. Systematic Biology 56:741-752.

Bromham, L., and D. Penny. 2003. The modern molecular clock. Nature Reviews Genetics 4:216-224.

Brown, J. W., R. B. Payne, and D. P. Mindell. 2007. Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson et al. Biology Letters 3:257-259.

Brown, J. W., J. S. Rest, J. García-Moreno, M. D. Sorenson, and D. P. Mindell. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biology 6:6.

Chiappe, L. M. 1995. The first 85 million years of avian evolution. Nature 378:349-355.

Chiappe, L. M., and G. J. Dyke. 2002. The Mesozoic radiation of birds. Annual Review of Ecology and Systematics 33:91-124.

Chiappe, L. M., and G. J. Dyke. 2006. The early evolutionary history of birds. Journal of the Palaontological Society of Korea 22:133-151.

Clarke, J. A., C. P. Tambussi, J. I. Noriega, G. M. Erikson, and R. A. Ketcham. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. Nature 433:305-308.

Cooper, A., and D. Penny. 1997. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. Science 275:1109-1113.

Cope, E. D. 1867. An account of the extinct reptiles which approached the birds. Proceedings of the Academy of Natural Sciences of Philadelphia 1867:234-235.

Cubo, J. 2003. Evidence for speciational change in the evolution of ratites (Aves : Palaeognathae). Biological Journal of the Linnean Society 80:99-106.

Cutler, D. J. 2000. Estimating divergence times in the presence of an overdispersed molecular clock. Molecular Biology and Evolution 17:1647–1660.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4:e88.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Dyke, G. J., and M. van Tuinen. 2004. The evolutionary radiation of modern birds (Neornithes): reconciling molecules, morphology and the fossil record. Zoological Journal of the Linnean Society 141:153-177.

Easteal, S. 1999. Molecular evidence for the early divergence of placental mammals. BioEssays 21:1052-1058.

Ericson, P. G. P., C. L. Anderson, T. Britton, A. Elzanowski, U. S. Johansson, M. Källersjö, J. I. Ohlson, T. J. Parsons, D. Zuccon, and G. Mayr. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biology Letters 4:543-547.

Feduccia, A. 1995. Explosive evolution in tertiary birds and mammals. Science 267:637-638.

Feduccia, A. 2003. 'Big bang' for tertiary birds? Trends in Ecology & Evolution 18:172-176.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368-376.

Friant, M. 1968. Sur les ceintures des membres des Oiseaux. Acta Anatomica 69:262-273.

Galtier, N., and M. Gouy. 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Molecular Biology and Evolution 15:871-879.

Gillespie, J. H. 1984. The molecular clock may be an episodic clock. Proceedings of the National Academy of Sciences of the United States of America 81:8009-8013.

Gillespie, J. H., and C. H. Langley. 1979. Are evolutionary rates really variable? Journal of Molecular Evolution 13:27-34.

Groth, J. G., and G. F. Barrowclough. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Molecular Phylogenetics and Evolution 12:115-123.

Hackett, S. J., R. T. Kimball, S. Reddy, R. C. K. Bowie, E. L. Braun, M. J. Braun, J. L. Chojnowski, W. A. Cox, K.-L. Han, J. Harshman, C. J. Huddleston, B. D. Marks, K. J. Miglia, W. S. Moore, F. H. Sheldon, D. W. Steadman, C. C. Witt, and T. Yuri.

2008. A phylogenomic study of birds reveals their evolutionary history. Science 320:1763-1768.

Haddrath, O., and A. J. Baker. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. Proceedings of the Royal Society of London B Biological Sciences 268:939-945.

Harrison, G. L., P. A. McLenachan, M. J. Phillips, K. E. Slack, A. Cooper, and D. Penny. 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late Cretaceous. Molecular Biology and Evolution 21:974-983.

Hedges, S. B., and S. Kumar (eds) 2009. The Timetree of Life. Oxford University Press, New York.

Hedges, S. B., P. H. Parker, C. G. Sibley, and S. Kumar. 1996. Continental breakup and the ordinal diversification of birds and mammals. Nature 381:226-229.

Heilman, G. 1926. The Origin of Birds. Witherby, London.

Hennig, W. 1966. Phylogenetic Systematics. University of Illinois Press, Urbana.

Himmelmann, L., and D. Metzler. 2009. TreeTime: an extensible C++ software package for Bayesian phylogeny reconstruction with time-calibration. Bioinformatics 25:2440-2441.

Hipsley, C., L. Himmelmann, D. Metzler, and J. Muller. 2009. Integration of Bayesian molecular clock methods and fossil-based soft bounds reveals early Cenozoic origin of African lacertid lizards. BMC Evolutionary Biology 9:151.

Ho, S. Y. W. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. Journal of Avian Biology 38:409-414.

Ho, S. Y. W., and G. Larson. 2006. Molecular clocks: when times are a-changin'. Trends in Genetics 22:79-83.

Ho, S. Y. W., and M. J. Phillips. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Systematic Biology 58:367-380.

Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Molecular Biology and Evolution 22:1561-1568.

Holmgren, N. 1955. Studies on the phylogeny of birds. Acta Zoologica 36:243-328.

Huelsenbeck, J. P., B. Larget, and D. L. Swofford. 2000. A compound Poisson process for relaxing the molecular clock. Genetics 154:1979-1892.

Hug, L. A., and A. J. Roger. 2007. The impact of fossils and taxon sampling on ancient molecular dating analyses. Molecular Biology and Evolution 24:1889–1897.

Huxley, T. H. 1868. On the animals which are most nearly intermediate between birds and reptiles. Annals and Magazine of Natural History 4:66-75.

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

Kolaczkowski, B., and J. W. Thornton. 2008. A mixed branch length model of heterotachy improves phylogenetic accuracy. Molecular Biology and Evolution 25:1054-1066.

Ksepka, D. T. 2009. Broken gears in the avian molecular clock: new phylogenetic analyses support stem galliform status for Gallinuloides wyomingensis and rallid affinities for Amitabha urbsinterdictensis. Cladistics 25:173-197.

Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. Nature 392:917-920.

Lartillot, N., and H. Philippe. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Molecular Biology and Evolution 21:1095-1109.

Lepage, T., D. Bryant, H. Philippe, and N. Lartillot. 2007. A general comparison of relaxed molecular clock models. Molecular Biology and Evolution 24:2669-2680.

Linder, H. P., C. R. Hardy, and F. Rutschmann. 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. Molecular Phylogenetics and Evolution 35:569-582.

Lindsay, B. 1885. On the avian sternum. Proceedings of the Zoological Society of London 53:684-716.

Liu, L., and D. K. Pearl. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Systematic Biology 56:504-514.

Lopez, P., D. Casane, and H. Philippe. 2002. Heterotachy, an important process of protein evolution. Molecular Biology and Evolution 19:1-7.

Lowe, P. R. 1928. Studies and observations bearing on the phylogeny of the ostrich and its allies. Proceedings of the Zoological Society of London 16:185-247.

Lucas, F. A. 1916. On the beginnings of flight. American Museum Journal 16:5-11.

Magallón, S. A. 2004. Dating lineages: molecular and paleontological approaches to the temporal framework of clades. International Journal of Plant Sciences 165:S7-S21.

Margoliash, E. 1963. Primary structure and evolution of cytochrome c. Proceedings of the National Academy of Sciences of the United States of America 50:672-679.

Meise, W. 1963. Verhalten der staussartigen Vögel und monophylie der Ratitae. Proceedings of the XIII International Ornithological Congress 1:115-125.

Mivart, G. 1881. A popular account of chameleons. Nature 24:309-312.

Near, T. J., P. A. Meylan, and H. B. Shaffer. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. American Naturalist 165:137-146.

Near, T. J., and M. J. Sanderson. 2004. Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. Philosophical Transactions of the Royal Society of London B Biological Sciences 359:1477-1483.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53:47-67.

Pagel, M., and A. Meade. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. Systematic Biology 53:571-581.

Pagel, M., and A. Meade. 2008. Modelling heterotachy in phylogenetic inference by reversible-jump Markov chain Monte Carlo. Philosophical Transactions of the Royal Society of London B Biological Sciences 363:3955-3964.

Pagel, M., C. Venditti, and A. Meade. 2006. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science 314:119-121.

Paton, T., O. Haddrath, and A. J. Baker. 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proceedings of the Royal Society of London B Biological Sciences 269:839-846.

Pereira, S. L., and A. J. Baker. 2006a. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Molecular Biology and Evolution 23:1731-1740.

Pereira, S. L., and A. J. Baker. 2006b. A molecular timescale for galliform birds accounting for uncertainty in time estimates and heterogeneity of rates of DNA substitutions across lineages and sites. Molecular Phylogenetics and Evolution 38:499-509.

Pereira, S. L., and A. J. Baker. 2008. DNA evidence for a Paleocene origin of the Alcidae (Aves: Charadriiformes) in the Pacific and multiple dispersals across northern oceans. Molecular Phylogenetics and Evolution 46:430-445.

Pereira, S. L., K. P. Johnson, D. H. Clayton, and A. J. Baker. 2007. Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and a dispersal-driven radiation in the Paleogene. Systematic Biology 56:656 - 672.

Rannala, B., and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645-1656.

Rannala, B., and Z. Yang. 2007. Inferring speciation times under an episodic molecular clock. Systematic Biology 56:453-466.

Rodríguez, F., J. L. Oliver, A. Marín, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. Journal of Theoretical Biology 142:485-501.

Rutschmann, F. 2006. Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. Diversity and Distributions 12:35-48.

Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Molecular Biology and Evolution 14:1218-1231.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 19:101-109.

Sanderson, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301-302.

Sibley, C. G., and J. E. Ahlquist. 1990. Phylogeny and Classification of Birds. Yale University Press, London.

Slack, K. E., C. M. Jones, T. Ando, G. L. Harrison, R. E. Fordyce, and D. Penny. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Molecular Biology and Evolution 23:1144-1155.

Squartini, F., and P. F. Arndt. 2008. Quantifying the stationarity and time reversibility of the nucleotide substitution process. Molecular Biology and Evolution 25:2525-2535.

Steel, M. 2005. Should phylogenetic models be trying to 'fit an elephant'? Trends in Genetics 21:307-309.

Svennblad, B. 2008. Consistent estimation of divergence times in phylogenetic trees with local molecular clocks. Systematic Biology 57:947 - 954.

Thorne, J. L. 2003. MULTIDISTRIBUTE. Available from the author (http://statgen.ncsu.edu/thorne/multidivtime.html).

Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation. Systematic Biology 51:689-702.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of evolution. Molecular Biology and Evolution 15:1647-1657.

Tuffley, C., and M. Steel. 1998. Modeling the covarion hypothesis of nucleotide substitution. Mathematical Biosciences 147:63-91.

van Tuinen, M. 2009. Birds (Aves). Pages 409-411 *in* The Timetree of Life (S. B. Hedges, and S. Kumar, eds.). Oxford University Press.

van Tuinen, M., and G. J. Dyke. 2004. Calibration of galliform molecular clocks using multiple fossils and genetic partitions. Molecular Phylogenetics and Evolution 30:74-86.

van Tuinen, M., and E. A. Hadly. 2004a. Calibration and error in placental molecular clocks: a conservative approach using the Cetartiodactyl fossil record. Journal of Heredity 95:200-208.

van Tuinen, M., and E. A. Hadly. 2004b. Error in estimation of rate and time inferred from the early amniote fossil record and avian molecular clocks. Journal of Molecular Evolution 59:267-276.

van Tuinen, M., and S. B. Hedges. 2001. Calibration of avian molecular clocks. Molecular Biology and Evolution 18:206-213.

van Tuinen, M., C. G. Sibley, and S. B. Hedges. 2000. The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. Molecular Biology and Evolution 17:451-457.

van Tuinen, M., T. A. Stidham, and E. A. Hadly. 2006. Tempo and mode of modern bird evolution observed with large-scale taxonomic sampling. Historical Biology 18:205-221.

Venditti, C., A. Meade, and M. Pagel. 2006. Detecting the node-density artifact in phylogeny reconstruction. Systematic Biology 55:637-643.

Verheyen, R. A. 1960. Les nandous (Rheiformes) sont apparentes aux tinamous (Tinamidae/Galliformes). Le Gerfaut 50:289-293.

Vogt, C. 1880. Archaeopteryx macroura, an intermediate form between birds and reptiles. Ibis 22:434-456.

von Blotzheim, U. 1960. Zur morphologie und ontogenese von schultergürtel, sternum und becken von Struthio, Rhea und Dromiceius: ein beitrag zur phylogenese der Ratiten. Proceedings of the 7th International Ornithological Congress 1:240-251.

von Steinmann, G. 1922. Laufvögel und flügvögel. Anatomische Anzeiger 55:239-244.

Waddell, P. J., Y. Cao, M. Hasegawa, and D. P. Mindell. 1999. Assessing the Cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. Systematic Biology 48:119-137.

Welch, J., E. Fontanillas, and L. Bromham. 2005. Molecular dates for the "Cambrian explosion": the influence of prior assumptions. Systematic Biology 54:672-678.

Welch, J. J., and L. Bromham. 2005. Molecular dating when rates vary. Trends in Ecology & Evolution 20:320-327.

Wiedersheim, R. 1884. Die stammesentwicklung der Vögel. Biologische Zentrallblatt 3:654-668; 688-695.

Wiedersheim, R. 1885. Ueber die vorfahren der heutigen Vögel. Humboldt 4:213-225.

Woese, C., L. Achenbach, P. Rouviere, and L. Mandelco. 1991. Archaeal phylogeny: reexamination of the phylogenetic position of Archaeoglobus fulgidus in light of certain composition-induced artifacts. Systematic and Applied Microbiology 14:364-371.

Wright, T. F., E. E. Schirtzinger, T. Matsumoto, J. R. Eberhard, G. R. Graves, J. J. Sanchez, S. Capelli, H. Muller, J. Scharpegge, G. K. Chambers, and R. C. Fleischer. 2008. A multilocus molecular phylogeny of the parrots (Psittaciformes): support for a Gondwanan origin during the Cretaceous. Molecular Biology and Evolution 25:2141-2156.

Wu, J., E. Susko, and A. J. Roger. 2008. An independent heterotachy model and its implications for phylogeny and divergence time estimation. Molecular Phylogenetics and Evolution 46:801-806.

Yang, Z. 1994. Estimating the pattern of nucleotide substitution. Journal of Molecular Evolution 39:105-111.

Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends in Ecology & Evolution 11:367-372.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24:1586-1591.

Yang, Z., and B. Rannala. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Molecular Biology and Evolution 23:212-226.

Yoder, A. D., and Z. Yang. 2000. Estimation of primate speciation dates using local molecular clocks. Molecular Biology and Evolution 17:1081-1090.

Zuckerkandl, E., and L. Pauling. 1962. Molecular disease, evolution, and genic heterogeneity. Pages 189-225 *in* Horizons in Biochemistry (M. Kasha, and B. Pullman, eds.). Academic Press, New York.

Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. Pages 97-166 *in* Evolving Genes and Proteins (V. Bryson, and H. J. Vogel, eds.). Academic Press, New York.

Chapter 2

Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson et al. (2006)[1]

The discrepancy between fossil- and molecular-based age estimates for the diversification of modern birds has persisted despite increasingly large datasets on both sides (Penny and Phillips, 2004). For the purpose of addressing this discrepancy, Ericson et al. (2006) recently generated a significant neoavian dataset that is well represented by taxa (87 species comprising 75 traditional families), characters (five nuclear genes) and fossil calibrations (n = 23). The divergence times reported in this study are by far the youngest yet reported from genetic data. These authors conclude that there is no reliable molecular support for extensive diversification of Neoaves in the Cretaceous. While an increased agreement with the fossil record is encouraging (and, indeed, sought after), we find a number of problems with their study that calls this conclusion into question.

Our first concern with this paper involves the particular fossils used to calibrate and constrain estimated divergence times. Fossils are of fundamental importance in estimating dates with molecular sequence data, and care should be taken that they are taxonomically and stratigraphically well identified. While the fossils used in Ericson et al. (2006) appear to fit these criteria, we nevertheless take issue with the particular fossils used. First, Ericson et al. (2006) use a stem group galliform fossil (53 million years (Ma); their calibration 'F'; Mayr and Weidig, 2004) to date the divergence between Galliformes and Anseriformes, despite the fact that an older (66 Ma), and therefore more informative, fossil anseriform calibration exists (Clarke et al., 2005). Ericson et al.'s (2006) estimate of the age of the Galliformes–Anseriformes split is approximately 53 Ma, 13 Ma younger than the minimum age definitively known from the fossil record (Benton and Donoghue,

---

[1] Published as Brown, J. W., R. B. Payne, and D. P. Mindell. 2007. Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson et al. Biology Letters 3:257-259.

2006). Second, for the fixed-age calibration required for their choice of analysis (i.e. a node whose age if assumed known without error), they use a 47.5 Ma stem group representative of Trochilidae to mark the splitting of hummingbirds from other Apodiformes (their calibration 'Q'). No rationale is given explaining why this particular fossil was adopted, and we note that an older (62 Ma), more derived and hence more appropriate fossil is established from the stem of Sphenisciformes (Slack et al., 2006). Regardless, owing to the importance of the single fixed constraint, alternatives should have been considered to investigate the influence of fossil choice. Third, the authors impose a maximum constraint of 95 Ma on the age of Neoaves, despite the fact that earlier dates have been published (e.g. van Tuinen and Hedges, 2001; Pereira and Baker, 2006). Fourth, one of their fossil calibrations (stem Strigiformes; their calibration 'F') is uninformative for dating purposes, as it is superseded by an equally old (55 Ma) but more derived fossil (stem Coliiformes; their calibration 'E'). Finally, an error in analysis is revealed in that Ericson et al. (2006) estimate the age of Pandionidae at approximately 29 Ma, despite the fact that a 37 Ma crown Pandionidae fossil (Harrison and Walker, 1976) was purportedly used as a minimum age constraint for this node in all dating analyses (the calibration 'G').

Our second concern involves the reliance on the program PATHd8 (Britton et al., 2007) for estimating lineage ages, a new method which can be understood as a smoothed local clock approach, where rates are smoothed between sister lineages. Ericson et al. (2006) also used the program r8s (Sanderson, 2003), but dismissed these results simply because these dates are older than those generated by PATHd8 (although the older r8s dates are consistent with previous molecular-generated dates). The inferred dates from r8s directly contradict their claim of an absence of neoavian diversification in the Cretaceous. Agreement with the fossil record, while satisfying in terms of congruence, is not a sufficient criterion to arbitrate between sets of dates generated by different methods. Rather, arbitration should rely upon the performance of methods on both empirical and simulated data, and PATHd8 has yet to be tested in this way. To compare their PATHd8 results with those from a well-vetted program, we reanalyzed the data of Ericson et al. (2006) using a Bayesian modelling of rate evolution (Thorne and Kishino, 2002) and the revised calibrations outlined previously (see *Methods*, below). Contrary to their results,

we find evidence for substantial diversification of Neoaves in the Cretaceous (Figure 2.1).

Finally, and most importantly, nowhere do Ericson et al. (2006) mention any error intervals on their dating estimates. Given the proximity of many nodes to the K–T boundary, confidence intervals on age estimates would cross into the Cretaceous and render their conclusion untenable. Error estimates are easily generated using either non-parametric bootstrapping or considering a posterior distribution of trees. As error is inherent in each step of molecular dating (sequences, alignment, fossils, trees, etc.), the lack of error calculation is disturbing and undermines their ultimate assertion. When incorporating error intervals in our reanalysis, 24 basal neoavian divergences are restricted to the Cretaceous (Figure 2.1, green bars). Of these, 15 lead directly to extant families. While the addition of further family representatives will undoubtedly break up some of these branches (forming crown clades), a Tertiary origin for much of Neoaves is clearly rejected. Given the results of our reanalysis of the data of Ericson et al. (2006), the noteworthy problems attendant in their study, and the plurality of genetic studies indicating a Cretaceous origin of modern birds, we respectfully disagree with their conclusion and find instead that there is no reliable molecular evidence against an extensive pre-Tertiary radiation of Neoaves.

ACKNOWLEDGEMENTS

I thank Ericson et al. for making their data freely available. I also thank I. Pop, R. Asheton, S. Asheton and D. Alexander for their encouragement during this study.

MATERIALS AND METHODS[1]

Sequence data

We used the same alignment as was analyzed in Ericson et al. (2006), which was made freely available by those authors (http://www.nrm.se/inenglish/researchandcollections/zoology/vertebratezoology/birds.4.4e32c81078a8d9249800014590.html). From the combined total alignment of 5299 base pairs (bp), 1536 were deemed ambiguously aligned and so were excluded, yielding a final combined data set of 3763 bp (Table 2.1). Although data matrices differ notably in size across our two studies, we find no influence of our relatively conservative site exclusion approach on inferred date estimates (see below). No attempt was made at phylogenetic reconstruction; rather, we used their inferred tree (their Figure ESM-9) for dating purposes to make the results directly comparable.

Divergence time estimation

Divergence times were estimated using the MULTIDISTRIBUTE package (Thorne, 2003). This is a Bayesian approach to modelling rate heterogeneity across a tree in an ancestor-descendent fashion (Thorne and Kishino, 2002; Thorne et al., 1998). Estimates of the transition/transversion rate ratio κ and the gamma site class-specific rates under the F84+G model were calculated for each gene individually in the baseml program of the PAML 3.15 package (Yang, 1997). The output from baseml was used as the input for the MULTIDISTRIBUTE program estbranches, which produces maximum likelihood (ML) estimates of branch lengths and their approximate variance–covariance matrix for each gene. Finally, substitution rates and divergence times were estimated through MCMC approximation in Multidivtime. Here, the logarithm of the substitution rate at the end of a branch is modelled with a normal distribution, the mean of which has an expected value equal to the rate at the beginning of the branch. While rates are implicitly assumed to be autocorrelated from ancestor to descendent nodes, this autocorrelation may decay with

[1] Originally published as online supplementary information for Brown, J. W., R. B. Payne, and D. P. Mindell. 2007. Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson et al. Biology Letters 3:257-259.

increasing branch lengths.

The MULTIDISTRIBUTE package prunes the outgroup for this final analysis, so the taxa involved in dating are all representatives of Neognathae. We defined a diffuse prior for the age of the root of this tree: mean (rttm) = 100, standard deviation (rttmsd) = 40. To investigate the influence of this prior we also ran analyses assuming a much younger divergence: rttm = 80, rttmsd = 50. Additional priors, determined from an average across genes, are given in Table 2.2. 'Bigtime', the maximum age allowed for the root, was set at 160 Ma, as this is well beyond molecular estimates of the age of the divergence between Galloanseres and Neoaves. The 22 internal fossil constraints used are given in Table 2.3. The program was run without the assumption of correlated changes in substitution rates across genes. Following a burnin of 10^5 samples, 10^4 samples were taken at a sampling interval of 10^2. Analyses were repeated with different initial conditions to check for convergence of the MCMC chain. Results from analyses assuming our alternative prior distributions (Table 2.2) were indistinguishable. Chronograms were constructed using FigTree v1.0 (Rambaut, 2006) and TS-Creator (http://www.stratigraphy.org).

Influence of site exclusions, fossil constraints, and dating method

Given the numerous concerns we identify with the study of Ericson et al. (2006), it is interesting to examine which aspects of our reanalysis contribute to the discordance in inferred dates between our two studies. To this end, we also reanalyze the data of Ericson et al. (2006) in PATHd8 (Britton et al., 2007) using our alignment (Table 2.1) and different fossil complements. To make results comparable, we use the same topology as above. As PATHd8 cannot accommodate multiple genes, we estimated maximum likelihood branch lengths on the concatenated data matrix using the optimal DNA substitution model (TIM+I+G) as inferred using AIC in Modeltest (Posada and Crandall, 1998) and PAUP* (Swofford, 2003). We will refer to this tree hereafter as T_{ML}.

To determine whether our relatively conservative site exclusion approach contributed to the older dates inferred here, we analyzed T_{ML} in PATHd8 using the same settings as in Ericson et al. (2006). Specifically, we placed an upper limit of 95 Ma on the age of

Neoaves and fixed the age of divergence between Trochilidae (hummingbirds) and Apodidae (swifts) at 47.5 Ma. All other fossil calibrations (Table 3.3) were treated as minimum age constraints. We exclude the fossil representing stem Strigiformes (owls; constraint 'F' in Ericson et al., 2006) from all analyses as it was found to be superseded by the fossil from stem Coliiformes (mousebirds; constraint 'E' in Ericson et al. 2006). The resulting chronogram (Figure 2.2, green dashed lines) is in very close agreement with the published chronogram of Ericson et al. (2006). In fact, regarding the K-T boundary, analysis of our alignment yielded slightly younger dates with only two neoavian divergence point estimates lying in the Cretaceous (versus five in Ericson et al., 2006). We can therefore be confident that our conservative alignment is not responsible for the older dates presented in Figure 2.1. We note, however, that analysis of our alignment yielded reasonable age estimates for the divergence between Paleognathae and Neognathae and between paleognath families Rheidae and Apterygidae (Figure 2.2), contra to the results of Ericson et al. (2006). This may have come about because the less conservative data exclusion of Ericson et al. (2006) included ambiguously aligned sites involving these taxa.

We next analyzed T_{ML} in PATHd8 using the calibration recommendations outlined in our manuscript. Specifically, we set a liberal upper limit on the age of Neoaves at 120 Ma, and for the required fixed calibration used the fossil from stem Spheniscifformes (penguins) at 62 Ma (Slack et al., 2006). Additional fossils were again treated as minimum age constraints. Replaced fossil constraints are denoted by a prime (′) symbol in Table 3.3. The resulting chronogram (Figure 2.2, solid black lines) illustrates the strong systematic influence of our calibration changes on the inferred age estimates. In contrast to the estimates inferred using the original constraints of Ericson et al. (2006), substantial diversification is inferred to have occurred in the Cretaceous using our calibrations. For example, whereas we infer the initial divergence within Neoaves to be ~ 70 Ma using the original Ericson et al. (2006) calibrations, the estimate increases to ~ 94 Ma using our revised calibrations. Furthermore, ages for all neoavian divergences are on average 18.95 Ma older using our calibration scheme (range 0 – 41.5 Ma). Most striking, however, are the differences in age estimates for the required fixed nodes. While Ericson et al. (2006) fix the divergence between Trochilidae and Apodidae at 47.5 Ma, we

estimate this divergence at 86.5 Ma. Our older estimate of 86.5 Ma is similar to the 75.7 ± 7.8 Ma estimate reported by van Tuinen and Hedges (2001). In our PATHd8 analysis we fixed the origin of stem Sphenisciformes at 62 Ma, whereas Ericson et al. (2006) estimate this event at 55 Ma (their minimum age constraint for that node), which is 7 Ma younger than the minimum age known from the fossil record. Again, we find much younger dates for nodes associated with the paleognath outgroups than was found in Ericson et al. (2006).

Finally, we can compare age estimates generated in Multidivtime (Figure 2.1) and PATHd8 (Figure 2.2) using the same revised fossil constraints. Two important differences can be seen between these two figures. First, the initial divergence within Neoaves is estimated as slightly older at ~ 100 Ma using the Bayesian method of Multidivtime (versus ~ 94 Ma in PATHd8). We note that this finding was precluded by the maximum age constraint of 95 Ma set by Ericson et al. (2006). Second, the results from Multidivtime indicate a more gradual diversification of Neoaves than those from PATHd8. This is likely due to the ability of Multidivtime to accommodate information from individual genes, as these genes are likely informative in different parts of the tree. While Multidivtime tends to produce older age estimates than PATHd8 for many nodes, the difference is much slighter and insignificant when the large posterior credible intervals are included.

Table 2.1 Aligned nuclear DNA fragment lengths. The alignment is taken directly from Ericson et al. (2006). 'Original' refers to aligned sequence lengths prior to excluding ambiguously aligned nucleotide sites.

Gene	Original	This study	Excluded sites
β-fibrinogen (intron 7)	1705	944	24-110 121-123 126 127 168-170 176 199 200 214 232 240 241 253 256 257-275 304-315 362 363 366-375 377 378 395 401 402 420 429 432 440 451-453 468 493-514 532 541 561 596 609 612 613 619 660-675 679-684 690 720 746-748 750 770 771 774 777 778-784 793 794 797 812 813 826-829 843-1123 1144 1150 1151 1190-1192 1195-1203 1222 1224 1227 1228 1229 1231 1247-1251 1265 1271-1278 1294 1296 1307 1335-1340 1346 1350-1353 1359 1360 1389 1394 1396-1408 1455 1468-1472 1491-1509 1513-1529 1532-1540 1551-1572 1579-1656 1660 1663 1673-1685 1692-1701
c-*myc* (exon 3)	510	498	64-75
myoglobin (intron 2)	1061	720	24-30 34 42 54 77-88 91 109 110 118 133 134 137-140 160 172 182-339 347-349 362-364 373 374 385 392 393 408-418 455 485 548 565-568 591 597-600 604 605 618-621 633-637 663 675 697 701 721-723 735-751 759-801 832 839-846 874 875 983 1004-1023 1027 1036-1039
ornithine decarboxylase (ODC; introns 6,7; exon 7)	1093	671	80-82 86-178 184 192 193 198 208 213 215-276 287 289 292 299 300 307 333 339 346 354-356 364 373-386 391-399 404 411-427 438 448 450 451 545 546 558-563 588-648 658 670-683 690-692 702-707 712-726 734 735 742 753 762-770 781 787 790 791 799 808 812-822 828 832 833 839-860 866 869 877-879 892 913 918 926 927 943 948 952 953 956 974 975 984-988 994 999 1005-1008 1013 1014 1031-1034 1036 1042 1056
RAG-1	930	930	-
Total	5299	3763	1536

Table 2.2 Priors used in the Multidivtime analyses. All priors were determined from an average across the five genes. Those listed under 'Conventional Prior' were used to generate the results presented in Figure 1; these are the priors advocated by the authors of the method. To examine the influence of prior probabilities on inferred age estimates Multidivtime was also ran with deliberately young age priors (listed under 'Young Prior').

Parameter	Conventional Prior	Young Prior
rttm (mean time separating root and present)	100	80
rttmsd (standard deviation of rttm)	40	50
rtrate (mean rate at root node)	0.0016	0.0013
rtratesd (standard deviation of rtrate)	0.0016	0.0013
brownmean (mean of Brownian motion constant)	0.01875	0.015
brownmeansd (standard deviation of brownmean)	0.01875	0.015
bigtime (oldest age allowed for root node)	160	160

Table 2.3 Fossil calibrations used in this study. For fossils used in Ericson et al. (2006), see that paper for the original fossil references. All fossils are treated as minimum age constraints, and are placed in the tree exactly as in Ericson et al. (2006). Replaced fossil calibrations used in our reanalysis are denoted by a prime (′) symbol. Fossil calibration 'F' (stem group Strigiformes) from Ericson et al. (2006) was not used as it was found to be redundant with the more derived fossil calibration E (stem group Coliiformes).

Symbol	Fossil calibration	Age (Ma)	Source
A	crown Pici	30	Ericson et al. (2006)
B	stem Upupidae + Phoeniculidae	47.5	Ericson et al. (2006)
C	stem Coraciidae + Brachypteraciidae	47.5	Ericson et al. (2006)
D	stem Trogoniformes	53	Ericson et al. (2006)
E	stem Coliiformes	55	Ericson et al. (2006)
G	crown Pandionidae	37	Ericson et al. (2006)
H	stem Cariamidae	47.5	Ericson et al. (2006)
I	stem Phalacrocoracidae	25	Ericson et al. (2006)
J	crown Sulidae	33	Ericson et al. (2006)
K	stem Fregatidae	53	Ericson et al. (2006)
L	stem Sphenisciformes	55	Ericson et al. (2006)
L′	stem Sphenisciformes	62	Slack et al. (2006)
M	crown Balaenicipitidae	30	Ericson et al. (2006)
N	crown Heliornithidae	14	Ericson et al. (2006)
O	stem Jacanidae	30	Ericson et al. (2006)
P	stem Apodiformes	53	Ericson et al. (2006)
Q	stem Trochilidae	47.5	Ericson et al. (2006)
Q′	stem Trochilidae	30	Mayr (2004)
R	crown Pteroclididae	30	Ericson et al. (2006)
S	stem Phoenicopteriformes	30	Ericson et al. (2006)
T	stem Phaethontidae	55	Ericson et al. (2006)
U	stem Galliformes	53	Ericson et al. (2006)
U′	Stem Anatidae	66	Clarke et al. (2005)
V	stem Gruidae + Aramidae	30	Ericson et al. (2006)
X	Stem Gaviiformes	30	Ericson et al. (2006)

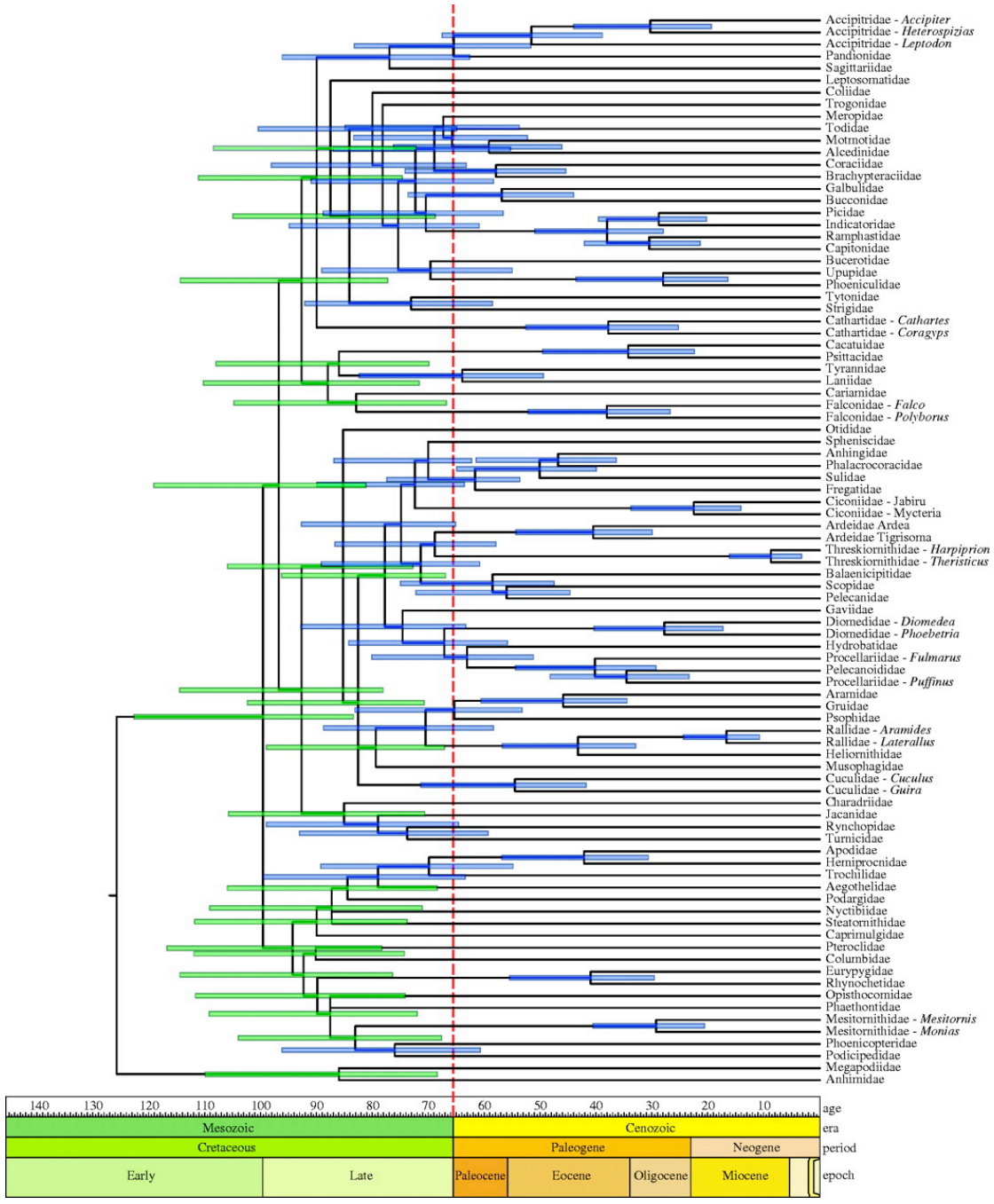

Figure 2.1 Chronogram for Neoaves estimated using a Bayesian modelling of rate evolution. The dashed vertical red line marks the K–T boundary. Error bars represent posterior probability (0.95) credible intervals (root node 104-154 Ma). An unambiguous ancient diversification of Neoaves is indicated by 24 credible intervals restricted to the Cretaceous (green bars).

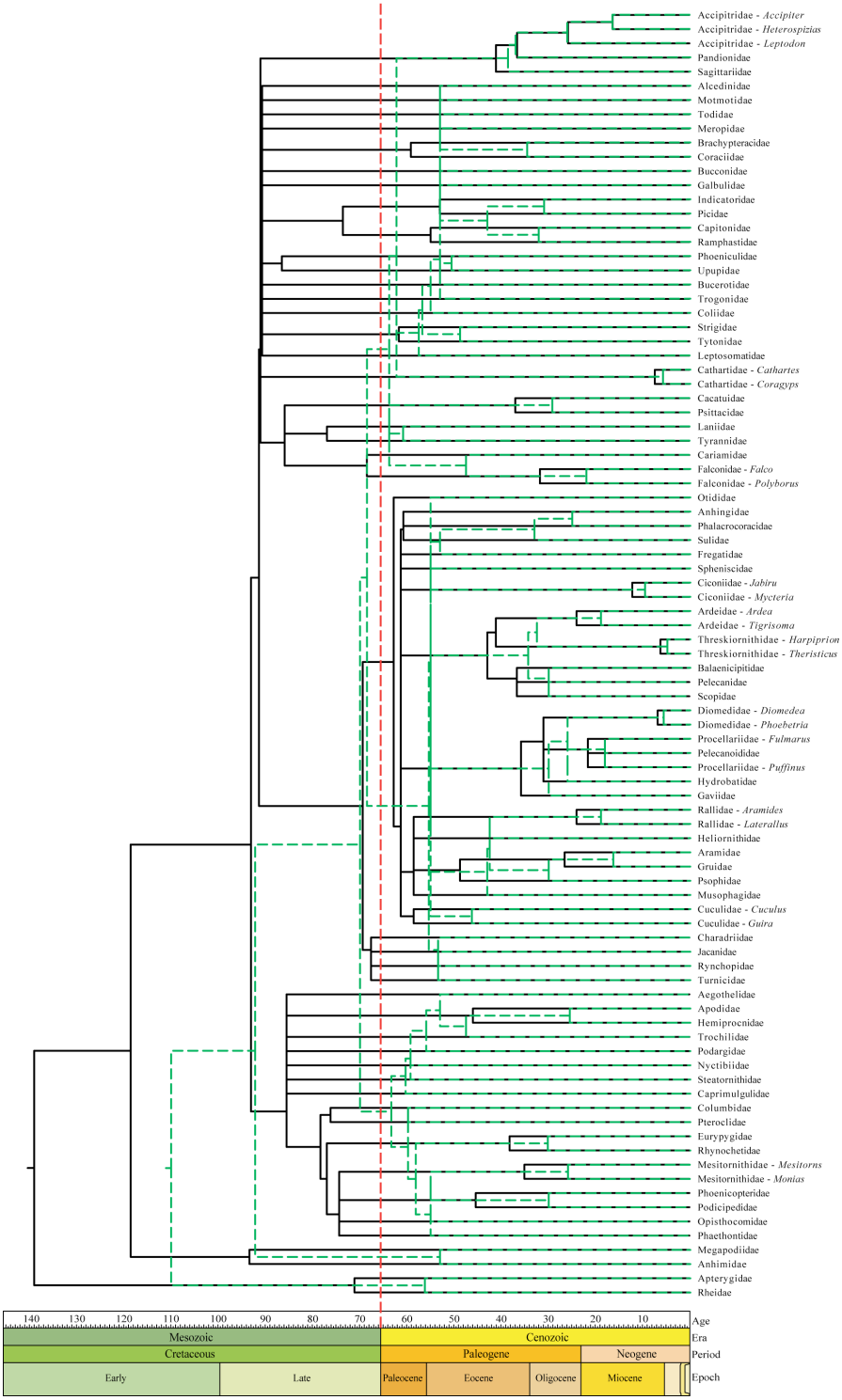

Figure 2.2 Chronograms generated using PATHd8 on the concatenated data matrix implementing the fossil constraints of Ericson et al. (2006; green dashed chronogram) and those of our reanalysis (black solid chronogram). The dashed vertical red line marks the K-T boundary.

REFERENCES

Benton, M. J. and P. C. J. Donoghue. 2006. Paleontological evidence to date the tree of life. Molecular Biology and Evolution 24:26-53.

Britton, T., C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer. 2007. Estimating divergence times in large phylogenetic trees. Systematic Biology 56:741-752.

Clarke, J. A., C. P. Tambussi, J. I. Noriega, G. M. Erikson, and R. A. Ketcham. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. Nature 433:305-308.

Ericson, P. G. P., C. L. Anderson, T. Britton, A. Elzanowski, U. S. Johansson, M. Källersjö, J. I. Ohlson, T. J. Parsons, D. Zuccon, and G. Mayr. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biology Letters 4:543-547.

Harrison, C. J. O., and C. A. Walker. 1976. Birds of the British Upper Eocene. Zoological Journal of the Linnean Society 59:323-351.

Mayr, G. 2004. Old world fossil record of modern-type hummingbirds. Science 304:861-864.

Mayr, G. and I. Weidig. 2004. The Early Eocene bird Gallinuloides wyomingensis—a stem group representative of Galliformes. Acta Palaeontologica Polonica 49:211-217.

Penny, D. and M. J. Phillips. 2004. The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution?. Trends in Ecology and Evolution 19:516-522.

Pereira, S. L. and A. J. Baker. 2006. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Molecular Biology and Evolution 23:1731-1740.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817-818.

Rambaut, A. 2006. FigTree v1.0. Available from the author (http://evolve.zoo.ox.ac.uk/software.html?id=figtree).

Sanderson, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302.

Slack, K. E., C. M. Jones, T. Ando, G. L. Harrison, R. E. Fordyce, and D. Penny. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Molecular Biology and Evolution 23:1144-1155.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). version 4 (beta 10). Sinauer Associates, Sunderland, Massachusetts.

Thorne, J. L. 2003. MULTIDISTRIBUTE. Available from the author (http://statgen.ncsu.edu/thorne/multidivtime.html).

Thorne, J. L. and H. Kishino. 2002. Divergence time and evolutionary rate estimation. Systematic Biology 51:689-702.

Thorne, J. L, H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of evolution. Molecular Biology and Evolution 15:1647-1657.

van Tuinen, M. and S. B. Hedges. 2001. Calibration of avian molecular clocks. Molecular Biology and Evolution 18:206-213.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in BioSciences 13:555-556.

Chapter 3

Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages[1]

ABSTRACT

Background: Determining an absolute timescale for avian evolutionary history has proven contentious. The two sources of information available, paleontological data and inference from extant molecular genetic sequences (colloquially, 'rocks' and 'clocks'), have appeared irreconcilable; the fossil record supports a Cenozoic origin for most modern lineages, whereas molecular genetic estimates suggest that these same lineages originated deep within the Cretaceous and survived the K-Pg (Cretaceous-Paleogene; formerly Cretaceous-Tertiary or K-T) mass-extinction event. These two sources of data therefore appear to support fundamentally different models of avian evolution. The paradox has been speculated to reflect deficiencies in the fossil record, unrecognized biases in the treatment of genetic data or both. Here we attempt to explore uncertainty and limit bias entering into molecular divergence time estimates through: (i) improved taxon (n = 135) and character (n = 4594 bp mtDNA) sampling; (ii) inclusion of multiple cladistically tested internal fossil calibration points (n = 18); (iii) correction for lineage-specific rate heterogeneity using a variety of methods (n = 5); (iv) accommodation of uncertainty in tree topology; and (v) testing for possible effects of episodic evolution.

Results: The various 'relaxed clock' methods all indicate that the major (basal) lineages of modern birds originated deep within the Cretaceous, although temporal intraordinal diversification patterns differ across methods. We find that topological uncertainty had a systematic but minor influence on date estimates for the origins of major clades, and Bayesian analyses assuming fixed topologies deliver similar results to analyses with

---

[1] Published as Brown, J. W., J. S. Rest, J. García-Moreno, M. D. Sorenson, and D. P. Mindell. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biology 6:6.

unconstrained topologies. We also find that, contrary to expectation, rates of substitution are not autocorrelated across the tree in an ancestor-descendent fashion. Finally, we find no signature of episodic molecular evolution related to either speciation events or the K-Pg boundary that could systematically mislead inferences from genetic data.

Conclusion: The 'rock-clock' gap has been interpreted by some to be a result of the vagaries of molecular genetic divergence time estimates. However, despite measures to explore different forms of uncertainty in several key parameters, we fail to reconcile molecular genetic divergence time estimates with dates taken from the fossil record; instead, we find strong support for an ancient origin of modern bird lineages, with many extant orders and families arising in the mid-Cretaceous, consistent with previous molecular estimates. Although there is ample room for improvement on both sides of the 'rock-clock' divide (e.g. accounting for 'ghost' lineages in the fossil record and developing more realistic models of rate evolution for molecular genetic sequences), the consistent and conspicuous disagreement between these two sources of data more likely reflects a genuine difference between estimated ages of (i) stem-group origins and (ii) crown-group morphological diversifications, respectively. Further progress on this problem will benefit from greater communication between paleontologists and molecular phylogeneticists in accounting for error in avian lineage age estimates.

BACKGROUND

Many evolutionary models (Archibald and Deutschman, 2001; Cooper and Fortey, 1998; Penny and Phillips, 2004; van Tuinen et al., 2006) are tightly linked to absolute timescales. A reliable temporal framework is therefore required for understanding the tempo (and, through correlation with geophysical phenomena, mechanisms) of biological evolution. There are two complementary sources of information for dating ancient biological divergences: (1) physical historical remains (either paleontological or ichnological); and (2) molecular sequence differences among extant taxa, the analysis of which requires assumptions about the processes and rates of sequence evolution. Unfortunately, these two sources of information ('rocks' and 'clocks', respectively) often yield starkly disparate estimates of the timing of major biological divergences (Benton and Ayala, 2003).

Of course, some discrepancy is expected, as these two sources of data (molecular age, MA; fossil age, FA) concern different stages during the process of cladogenesis ($\delta_{True\ MA\text{-}FA}$; Figure 3.1). As fossils document products of evolution, they necessarily post-date speciation events. The underestimation of speciation times from fossil data ($\delta_{Fossil\ error}$) can be partitioned into two components: (i) the interval, following speciation, required for diagnostic characters to evolve ($\delta_{Diagnostic\ character}$); and (ii) the time, following the evolution of diagnostic characters, realized for the deposition of a sampled fossil ($\delta_{Oldest\ fossil}$). Here, $\delta_{Diagnostic\ character}$ can be regarded as a fixed value (although different for every node), however $\delta_{Oldest\ fossil}$ can be reduced with subsequently older fossil finds. In contrast to fossils, molecular data instead reflect genetic divergence, which must predate speciation events because genetic lineages present in two newly evolved sister lineages coalesce (on average) $2N_e$ generations prior to speciation (Edwards and Beerli, 2000). However, the errors associated with molecular age estimates ($\delta_{Coalescence}$ and $\delta_{Clock\ error}$) are more complex than analogous fossil errors. For example, if no polymorphism exists for a particular locus at speciation, then inferred genetic divergence times based on that locus will actually post-date speciation, as no information exists to trace the underlying genealogy. Furthermore, molecular data may overestimate or underestimate the true speciation time because of stochastic errors associated with divergence time estimation ($\delta_{Clock\ error}$), and this uncertainty will increase as one extrapolates backwards through time, even with an infinite amount of data (Rannala and Yang, 2007). Regardless, for a given node with a minimum age constraint derived from the fossil record, the realized discrepancy between the two estimates ($\delta_{Realized\ MA\text{-}FA} = MA - FA = \delta_{Molecular\ error} + \delta_{Fossil\ error}$) will always be positive, and is a parameter that both paleontologists and molecular biologists are working to minimize (Benton and Ayala, 2003).

Strictly speaking, any molecular estimate that generates a positive value of $\delta_{Realized\ MA\text{-}FA}$ is consistent with the fossil record. It is instead the *magnitude* of $\delta_{Realized\ MA\text{-}FA}$ that suggests conflict, and distressingly large values sometimes exist. Conflict between the two sources of information is especially evident with respect to the ages of extant avian lineages (Neornithes). Based on a strict interpretation of the fossil record (i.e. insignificant $\delta_{Fossil\ error}$), Feduccia (Feduccia, 1995; Feduccia, 2003) proposed an explosive Cenozoic origin for most modern avian lineages, presumably a result of open

niches left by newly extinct non-avian dinosaurs and other taxa. Although a recent fossil find (Clarke et al., 2005) forces a minimum of five of the earliest Neornithes divergences into the late Cretaceous, the fossil record generally supports the view that most modern lineages originated in the Cenozoic (Bleiweiss, 1998; Chiappe and Dyke, 2002; Feduccia, 1995; Feduccia, 1999; Feduccia, 2003; Fountaine et al., 2005; Mayr, 2005). In contrast, molecular estimates all indicate that these same lineages are considerably older, sometimes as much as twice as old as analogous paleontological estimates (Baker et al., 2007; Brown et al., 2007; Cooper and Penny, 1997; Hedges et al., 1996; Kumar and Hedges, 1998; Paton et al., 2002; Paton et al., 2003; Pereira and Baker, 2006; Rest et al., 2003; van Tuinen and Hedges, 2001; van Tuinen et al., 2006). Between these two extremes lies the Cretaceous-Paleogene (K-Pg; formerly Cretaceous-Tertiary or K-T) boundary, a period when as many as 50% of land-dwelling species went extinct (Benton, 1997). The conflicting age estimates thus have different implications regarding the influence of the K-Pg mass extinctions on the evolutionary radiation of modern birds.

Although resolution of this conflict is clearly important for understanding avian diversification, it is hindered by compelling arguments from both sides. The supposition that the quality of the fossil record deteriorates backwards in time, for example, is contradicted by the observed congruence between stratigraphic and phylogenetic ordering of taxa (Benton et al., 2000). Sophisticated stratigraphic analyses indicate that fossils of the antiquity necessary to produce congruence with molecular studies are extremely improbable (Bleiweiss, 1998; Foote et al., 1999; Foote and Sepkoski Jr., 1999) (but see (Tavaré et al., 2002; Wills, 2007)). Furthermore, of the known Mesozoic avian fossils (Benton, 1999; Chiappe and Dyke, 2002; Fountaine et al., 2005; Padian and Chiappe, 1998), the vast majority unambiguously lay outside Neornithes (Clarke and Chiappe, 2001). Although a few Cretaceous fossils putatively represent modern lineages (e.g. parrot (Stidham, 1998), loon (Chatterjee, 2002) and others (Chiappe and Dyke, 2002; Fountaine et al., 2005)) these have largely been dismissed as fragmentary and inconclusive (Chiappe and Dyke, 2002; Dyke and Mayr, 1999; Dyke and van Tuinen, 2004; Feduccia, 2003). One the molecular side, the recognition that rates of molecular evolution are often not clock-like (including birds (García-Moreno, 2004; Lovette, 2004; Mindell et al., 1996; Pereira and Baker, 2006)), and that lineage-specific heterogeneity is

common (Bromham and Penny, 2003), has spurred the development of numerous 'relaxed' molecular clock methods (see reviews in (Magallón, 2004; Rutschmann, 2006; Welch and Bromham, 2005)). In support of molecular genetic data, these methods perform well in simulation (Drummond et al., 2006; Ho et al., 2005) and, when applied to empirical data, deliver Cretaceous ages for the origin of modern birds (Brown et al., 2007; Pereira and Baker, 2006).

Given these arguments, the paleontological and molecular phylogenetic communities are currently at an impasse regarding the application of an absolute temporal axis for early organismal evolution (Benton, 1999; Easteal, 1999), and a range of evolutionary models (Archibald and Deutschman, 2001; Cooper and Fortey, 1998; Penny and Phillips, 2004; van Tuinen et al., 2006) remain viable for birds. Here we endeavour to determine whether a more rigorous treatment of molecular genetic data lessens the 'rock-clock' discrepancy ($\delta_{Realized\ MA\text{-}FA}$). In particular, we incorporate large fossil and taxon data sets, two components of molecular dating that have been shown to have a strong impact on the resulting divergence time estimates (Hug and Roger, 2007; Linder et al., 2005). In addition, we accommodate and explore the impact of uncertainty in both tree topology and molecular dating strategy. Finally, we test for signals of episodic molecular evolution related to both speciation events and absolute geologic time, processes that could potentially mislead molecular-based age estimates by systematically inflating branch lengths within speciose clades (Pagel et al., 2006a).

RESULTS

Phylogenetic inference

Our optimal phylogenetic reconstruction (T_{Optimal}; $\text{AIC}_c = 414160.2536$) is a significantly better fit to the mtDNA matrix than a recent consensus topology derived from the literature ($T_{\text{Consensus}}$; $\text{AIC}_c = 421460.9166$; see the methods section and Figure 3.2). Nevertheless, the two topologies agree in many instances. For example, several traditional orders identified as having little support for monophyly (e.g. Coraciiformes, Ciconiiformes, Caprimulgiformes and Falconiformes (Cracraft et al., 2004)) were also polyphyletic in our analyses. However, the two trees also differ in many respects, most

notably in the placement of Passeriformes. In $T_{\text{Consensus}}$, the clade is relatively derived in the tree, whereas in T_{Optimal} it forms the basal-most clade in Neoaves. Several traditionally hard-to-classify lineages (e.g. Pteroclidae, Opisthocomidae, Phaethontidae, Podargidae and Steatornithidae) are of suspect placement in T_{Optimal}. These and other uncertainties tend to be localized and do not (as we report below) overly influence date estimates for the basal nodes in the avian tree. Some of the topological differences, however, are of operational importance, as they cause either redundancy or obsolescence of some fossil constraints used in estimating divergence times. Overall, of the 18 total internal fossil calibrations considered, 16 were used on $T_{\text{Consensus}}$, and 17 on T_{Optimal} (Figure 3.2).

Divergence time estimation

A substantial signal was present for both a departure from a molecular clock and a lack of ancestor-descendant autocorrelation of substitution rates. Using penalized likelihood in r8s, both topologies $T_{\text{Consensus}}$ and T_{Optimal} were found to be unclock-like, with optimal smoothing values ($\log_{10}\lambda$) of 1.0 and 0.4, respectively. Analyses in Dating5 clearly rejected the constant-rate Poisson model (-ln L = 63214.8; χ^2 = 8051.61, df = 271, p = 0.000) but could not reject the stationary (high variance) rate model (-ln L = 2123.01; χ^2 = 268.352, df = 269, p = 0.482) which produced a large index of dispersion R = 469.782. Bayesian analyses in Multidivtime delivered positive but very small values for the degree of autocorrelation of substitution rates across both topologies (Table 3.1). Finally, analyses of $T_{\text{Consensus}}$ using BEAST indicated that non-autocorrelated models of rate variation fit the data significantly better than a molecular clock (Table 3.2). Of the non-autocorrelated models, the lognormal distribution (UCLN) had a much better harmonic mean model likelihood than the exponential distribution (UCED), and relaxation (T_{Flexible}) of a fixed topology further increased fit. Using each of these uncorrelated models, the covariance of substitution rates between ancestor and descendent branches across the tree was not significantly different from zero.

Given the extensive phylogenetic uncertainty within Neornithes, we focus on divergence times of clades for which monophyly is not contentious (Table 3.3). Two key trends are recognized. First, for a given dating method, differences in topology tended to have a minor but systematic influence on inferred ages. In general, T_{Optimal} delivered older

average date estimates than $T_{\text{Consensus}}$ using r8s (8.9 MY) and Multidivtime (3.6 MY), but the opposite trend was found with PATHd8 (-8.2 MY). When confidence/credible intervals are considered, however, topology did not significantly influence most individual date estimates. Overall, in terms of the degree of discrepancy between fossil and molecular dates on a whole-tree basis (average $\delta_{Realized\ MA\text{-}FA}$), topology had a noticeable (> 5 MY) influence on divergence estimates for only the PATHd8 analyses (Table 3.3).

Second, the choice of the relaxed clock method had a strong influence on inferred ages. R8s, Multidivtime and BEAST tended to deliver similar estimates for most clades of interest (Table 3.3). In contrast, PATHd8 generated considerably younger dates with much smaller confidence intervals, despite using the same bootstrapped phylograms and fossil constraints as r8s. Dating5 tended to produce the most extreme results, with inferred basal split estimates similar to those from Multidivtime, but some derived split estimates younger than those from PATHd8. Most significantly, PATHd8 and Dating5 together identified five of these major clades as having crown diversification restricted to the Cenozoic (Ratites, Charadriiformes, Procellariiformes, Cuculiformes and Apodiformes), although the remaining methods generate estimates for these same nodes that are on average more than 50% older. In terms of comparing molecular and fossil age estimates (average $\delta_{Realized\ MA\text{-}FA}$), r8s, Multidivtime and BEAST all show considerable discordance between the two sources of data, with the average molecular estimates for the major nodes (Table 3.3) being 36–45 MY older than corresponding fossil ages. PATHd8 and Dating5, in contrast, exhibit greater agreement between estimates from 'rocks' and 'clocks', with an average discrepancy of 17–25 MY.

Despite these differences, all methods agree that the basal splits within Neornithes occurred deep within the Cretaceous (Table 3.3, nodes A-E). The youngest estimate for the root of Neornithes (PATHd8, T_{Optimal}) is of Early Cretaceous age, 37 MY older than the oldest undisputed neornithean fossil (Clarke et al., 2005). Conflict among methods instead involves the diversification of derived lineages (Figures 2.3 and 2.4). Two patterns can be discerned. First, PATHd8 and Dating5 support bursts of speciation (many lineages arising almost simultaneously), whereas the remaining methods generally

support more gradual diversification. Second, and more germane to the 'rock-clock' problem, PATHd8 alone supports an extensive post-K-Pg radiation of Neoaves. For example, from results of the non-autocorrelated rate models in BEAST allowing topological uncertainty (T_{Flexible}; see Figure 3.4), not only are the basal splits inferred to have occurred in the Cretaceous, but 37 credible intervals (green bars) do not overlap the K-Pg boundary. Finally, no support is shown for episodic evolution, either correlated with speciation events ((Pagel et al., 2006a); no effect) or an increase in inferred substitution rate either during early diversification or following the K-Pg mass extinction (Figure 2.5).

DISCUSSION

Phylogenetic inference

Whether using fossil or molecular data, a daunting impediment to divergence time estimation in birds is that resolution of many inter-ordinal phylogenetic relationships has proven obstinate, despite large data matrices with multiple character types (Cracraft et al., 2004). Although our reconstruction T_{Optimal} is overly optimistic in being fully resolved, it provides a useful alternative to the conservative $T_{\text{Consensus}}$ (Figure 3.2).

T_{Optimal} recovers several traditional orders as polyphyletic (Caprimulgiformes, Coraciiformes, Falconiformes, Ciconiiformes), consistent with expectations (Cracraft et al., 2004) (but see (Livezey and Zusi, 2007)). Although T_{Optimal} has caprimulgiform (nightbirds) families much more distantly related to one another than previous morphological (Mayr, 2002) and molecular (Barrowclough et al., 2006) investigations, differences in taxon sampling confounds direct comparison across studies. While Coraciiformes (kingfishers and relatives) is not found to be monophyletic, the two recovered sub-groupings both fall within a larger clade containing owls (Strigiformes), parrots (Psittaciformes) and woodpeckers and relatives (Piciformes). The monophyletic status of the order Falconiformes has received mixed support in previous analyses (Ericson et al., 2006; Fain and Houde, 2004; Griffiths, 1994; Livezey and Zusi, 2007; Mayr and Clarke, 2003; Sibley and Ahlquist, 1990). Placement of Falconidae in T_{Optimal} is suspect and likely stems from insufficient taxon sampling from this family (Gibb et al.,

2007). Regardless, no support was found for hypotheses uniting falconiform taxa with owls (Strigiformes) (Livezey and Zusi, 2007) or New World vultures (Cathartidae) with storks (Ciconiiformes) (Sibley and Ahlquist, 1990).

Several monotypic avian families have traditionally proved difficult to classify. The enigmatic hoatzin (Opisthocomidae) has variously been allied with at least four distantly related orders (Galliformes, Cuculiformes, Musophagiformes and Columbiformes; see (Sibley and Ahlquist, 1990; Sorenson et al., 2003)). We find here an alliance with doves (Columbiformes), similar to joint analyses of mitochondrial and nuclear DNA sequences (Sorenson et al., 2003). The taxonomically problematic sandgrouse (family Pteroclidae) has alternatively been considered a member of Charadriiformes (shorebirds and allies (de Juana, 1997; Sibley and Ahlquist, 1990)) or Columbiformes (Ericson et al., 2006; Livezey and Zusi, 2007; Mayr and Clarke, 2003). Our reconstruction has the sandgrouse distantly related to both orders, and instead allied with Falconiformes. This relationship is unsupported elsewhere and we have little confidence in this placement. The novel placement of the tropicbird (family Phaethontidae) as sister to Sphenisciformes is similarly suspect.

Finally, we find no support in our mtDNA analyses for the neoavian clades 'Metaves' and 'Coronaves' identified from nuclear β-fibrinogen intron analyses (Fain and Houde, 2004), although our constraint tree allowed for this possibility ($T_{\text{Constraint}}$; see Figure 3.7). A major difference between these trees involves the phylogenetic position of the perching birds (Passeriformes); while nuclear DNA analyses recover Passeriformes as a relatively derived clade within 'Coronaves' (Ericson et al., 2006; Fain and Houde, 2004), in T_{Optimal} they instead comprise the basal-most lineage of Neoaves. This may be indicative of different phylogenetic signals in nuclear versus mtDNA sequences, as other mtDNA studies have also inferred a basal phylogenetic position of Passeriformes in Neoaves (Slack et al., 2007).

Uncertainty in tree topology and substitution rate evolution

While T_{Optimal} yields interesting hypotheses about avian relationships, the focus of this study is on estimating basal divergence times in Neornithes and we might regard

topology as a nuisance parameter (and explicitly so in the BEAST $T_{Flexible}$ analyses). Topological error is usually not considered in divergence time estimation, but potentially could systematically bias age estimates through: (i) incorrect placement of fossil calibrations; and (ii) improper estimation of branch lengths. Through our joint consideration of $T_{Consensus}$ and $T_{Optimal}$, we find that topology does have a systematic influence on inferred divergence times for nodes of interest (Table 3.3), but that for the present data set this influence differed in direction across methods and was generally insignificant when confidence/credible intervals were considered. Removal of the restriction of a fixed topology in BEAST ($T_{Flexible}$) yielded age estimates similar to those from Multidivtime analyses assuming $T_{Optimal}$. Although yielding diffuse estimates, this 'relaxed topology' approach better reflects uncertainty in the underlying data.

An interesting result reported here is that rates of molecular evolution are found to be non-autocorrelated across the Neornithes tree (Tables 3.1 and 3.2), a result previously noted for virus and marsupial data sets (Drummond et al., 2006). An autocorrelation of rates would imply an inheritance of the trait 'rate of evolution'. This makes intuitive sense when considering that ancestor and descendant lineages are likely similar in body size, generation time, DNA repair efficiency, population size and other traits influencing rates of sequence evolution, and the most popular molecular dating methods available indeed implicitly assume that rates are autocorrelated across a tree (Sanderson, 2003; Thorne et al., 1998). However, even if 'rate of evolution' is heritable, nodes separated by long periods of time may accumulate sufficient rate variation that autocorrelation in this trait will break down. Alternatively, 'rate of evolution' may simply be more labile than we expect. Regardless, violation of an implicit autocorrelation assumption did not have significant effects on inferred dates for nodes of interest (Table 3.3). For example, r8s and Multidivtime, which each deal with rate variation in an ancestor-descendant fashion, deliver age estimates that match quite closely to those generated by BEAST, which does not make such an assumption.

Antiquity of basal avian lineages

All methods employed here agree that the basal divergences within Neornithes occurred in the Cretaceous (Table 3.3, nodes A-E), supporting the refutation of a Cenozoic origin

of modern lineages (Feduccia, 1995; Feduccia, 2003) mandated by the discovery of the 66 MY duck *Vegavis iaai* (Clarke et al., 2005), which minimally forces five basal divergences into the Cretaceous. Moreover, our results are not dependent on this oldest fossil calibration, as analyses in r8s, PATHd8 and Multidivtime without using the *Vegavis* constraint returned nearly identical results to those reported here (data not shown); indeed, we must paradoxically conclude that this oldest undisputed neornithean fossil was essentially uninformative in our molecular dating analyses. Given the consensus across 'relaxed clock' methods employing very different assumptions about how molecular substitution rate evolves, we regard an Early Cretaceous origin of Neornithes as robustly supported. This inferred Cretaceous origin, and consequent survival of several avian lineages across the K-Pg boundary (Robertson et al., 2004), is consistent with previous molecular studies (Baker et al., 2007; Brown et al., 2007; Cooper and Penny, 1997; Hedges et al., 1996; Kumar and Hedges, 1998; Paton et al., 2002; Paton et al., 2003; Pereira and Baker, 2006; Rest et al., 2003; van Tuinen and Hedges, 2001; van Tuinen et al., 2006) and is supported by historical biogeography reconstructions (Cracraft, 2001).

An explanation occasionally offered for the antiquity of molecular dates is that rates of sequence change may speed up during radiations (Benton, 1999), consistent with a basic tenet of punctuated equilibrium theory where most character change is concomitant with speciation (Eldredge and Gould, 1972), possibly resulting from stochastic effects during founder effect speciation (Mayr, 1954; Pagel et al., 2006a). If true, an elevated number of molecular substitutions recorded during diversification could erroneously be interpreted as a longer time span at a slower background rate, resulting in a systematic overestimation of divergence times for all nodes predating the radiation. Some evidence exists for a correlation between speciation and character evolution (Barraclough and Savolainen, 2001; Cubo, 2003; Mindell et al., 1989; Pagel et al., 2006a), although a study of island radiations failed to detect such an effect (Bromham and Woolfit, 2004). While punctuated molecular evolution may be less frequent in animals (18% of cases versus 44% and 60% in plants and fungi, respectively (Pagel et al., 2006a)), this effect is nevertheless a prime candidate to explain the strong and persistent discrepancy between molecular- and fossil-based divergence estimates. However, we find no signal for

punctuated (speciational) molecular evolution (Pagel et al., 2006a) in the present data set. In addition, we fail to detect an accelerated rate associated with either the K-Pg boundary or during the initial diversification of Neornithes (Figure 3.5). If rates of sequence change were strongly influenced by diversification, we would expect clear departures from the inferred mean standardized substitution rate (Aris-Brosou and Yang, 2003). Although Cenozoic rates tend to be more variable than Mesozoic (ancestral) rates, we find no evidence for an obvious acceleration in substitution rate associated with any of the major nodes for any genetic partition. Although these two approaches to identifying episodic evolution would ideally involve more comprehensive taxon sampling, if punctuated evolution were primarily responsible for inflating inferred molecular dates then we likely would have detected the effect with the present data matrix.

Rather than narrowing the formidable 'rock-clock' gap through application of methods designed to minimize biases and accommodate uncertainty, we have instead considerably reinforced it. In part, the discordant age estimates can be explained through reference to the genuine time-lag ($\delta_{True\ MA\text{-}FA}$; see Figure 2.1) between the divergence of genetic lineages (predating speciation) and the evolution of diagnostic morphological characters (postdating speciation). However, given the estimated magnitude of $\delta_{Realized\ MA\text{-}FA}$ (Table 3.3), it is unlikely that $\delta_{True\ MA\text{-}FA}$ alone explains the dissonance. One the one hand, while the fossil record may be adequate for many evolutionary questions (Benton et al., 2000), there are clear instances where it is not. The 66 MY *Vegavis iaai* (Clarke et al., 2005), for example, requires the coexistence of Paleognathae representatives; however, Cretaceous paleognaths are unknown. This may be a result of a geographical bias in fossil sampling favouring the northern hemisphere (Cooper and Fortey, 1998; Cooper and Penny, 1997; Cracraft, 2001; Dyke, 2001; Hope, 2002; Smith and Peterson, 2002), as many taxa are hypothesized as having southern hemisphere (Gondwana) origins (Cracraft, 2001). Although fossil gap analysis implies that a Cretaceous origin of several avian orders is unlikely (Bleiweiss, 1998), this method improperly assumes that fossils are randomly distributed (uniformly recovered through time) since the origin of a taxon; alternative fossil recovery curves can support very different scenarios, including scenarios that are more consistent with molecular genetic timelines (Marshall, 1999), even when the fossil record is particularly sparse (Tavaré et al., 2002). Although rightly considered with

caution, the increasing number of fragmentary remains of putative neornithean lineages from the Cretaceous (Hope, 2002) lends credence to the ancient origin and diversification of Neornithes. On the other hand, there may yet be some unrecognized biases in the analysis of molecular genetic sequences, and our results suggest new directions for future avian divergence time studies (described below).

Radiation of Neornithes

Despite broad agreement on the antiquity of basal divergences in Neornithes, arbitration among macroevolutionary models (Archibald and Deutschman, 2001; Cooper and Fortey, 1998; Penny and Phillips, 2004; van Tuinen et al., 2006) to best describe the history of more derived lineages is complicated by disparity amongst various 'relaxed clock' results. Two important points of distinction can be recognized (Figures 3.3 and 3.4). First, Dating5 (overdispersed clock) and PATHd8 (rate smoothing across sister lineages) both infer bursts of speciation across the avian tree, while r8s (smoothing in an ancestor-descendant fashion), Multidivtime (Bayesian modelling of ancestor-descendant autocorrelated rate evolution) and BEAST (Bayesian modelling of rate evolution without an autocorrelation assumption or fixed topology) instead infer a more gradual diversification of Neornithes. Second, PATHd8 alone supports a prominent radiation of avian families in the Cenozoic, a scenario statistically rejected in many instances by the remaining four analyses. Although published PATHd8 divergence time estimates for Neoaves using nuclear DNA produced similarly young estimates (Ericson et al., 2006), a reanalysis of these same data using Multidivtime roundly refuted the findings (Brown et al., 2007), echoing the incongruence of PATHd8 reported here. While the better reconciliation between molecular and fossil age estimates that PATHd8 offers seems satisfying at first, the unique discrepancy of this method from the other much more rigorous and biologically realistic methods raises concern.

Given the apparent lack of autocorrelation of substitution rates, together with the intrinsic topological uncertainty in the Neornithes tree, the analyses in BEAST best reflect our current understanding of early avian evolution (Figure 3.4). Without the restrictive assumptions inherent in other 'relaxed clock' methods, these analyses unambiguously support a Cretaceous origin and diversification of basal avian lineages.

Even when considering large inferred credible intervals, 37 early avian divergences are restricted to the Cretaceous, similar to that found through the analysis of nuclear DNA sequences (Brown et al., 2007). It should be noted, however, that these nodes mostly represent order- and superfamily-level divergences; the majority of families sampled here (80 of 100 in BEAST) have their origins either overlapping or restricted to the Paleogene, consistent with interpretations from the fossil record (Dyke, 2001). In this respect, our results are similar to those inferred from a comprehensive study of the tempo of early mammalian evolution (Bininda-Emonds et al., 2007). The results of both studies indicate that significant cladogenesis occurred in the Cretaceous, but that many crown groups diversified in the Cenozoic.

Future progress

While there is a growing consensus for the Cretaceous origin of many Neornithes orders and families, the rate and timing of their diversification (and the influence of the K-Pg mass extinctions upon that diversification) is not yet resolved. MtDNA may have further utility in making progress on the problem, as our analysis of posterior credible interval widths indicates that longer sequences would likely improve divergence time estimates (Figure 3.6). However, we recognize that mtDNA is limited in that all sites belong to a single 'super-locus', and so a significant reduction of uncertainty (e.g. the slope in Figure 3.6) will ultimately require the supplement of multiple independent nuclear loci. In addition, while the present study was focused at the family level, increased taxon sampling will break up long branches (benefiting phylogenetic reconstruction) and improve branch length estimates. Nevertheless, our results suggest fertile ground for future molecular research into this problem. For example, we find that: (i) variation in the number of substitutions across branches can be accommodated by a high variance stationary-rate model (Cutler, 2000b); and (ii) rates are not autocorrelated across the avian tree in an ancestor-descendent fashion. This suggests potential for development of a hybrid model that accommodates both pieces of information and that future studies test assumptions of rate autocorrelation before invoking them, at least for the tree depth that we consider here.

In regards to the 'rock-clock' debate (Benton, 1999; Easteal, 1999), we feel that much

of the perceived dissonance between fossil- and molecular-based age estimates stems from a comparison of different phylogenetic entities: molecular phylogeneticists generally deal with stem-group origins, while paleontologists generally concentrate on crown-group diversification (van Tuinen et al., 2006). Moreover, it is too rarely emphasized that when dating the same node a genuine discrepancy is expected, as coalescent times (T_{gene}; see Figure 3.1) will predate cladogenesis while morphological diversification ($T_{morphology}$; see Figure 3.1) will postdate cladogenesis. The reality is that in normal practice neither group directly addresses the main parameter of interest, the timing of speciation ($T_{species}$; see Figure 3.1), the pattern of which is essential to the understanding of the origins of biodiversity. However, tools do exist to better estimate speciation times from both fossils (e.g. accounting for 'ghost' lineages (Tavaré et al., 2002)) and genetic data (e.g. explicitly modelling ancestral effective population sizes to account for coalescent times (Rannala and Yang, 2003)). Further, molecular methods can be augmented with greater information from the fossil record, possibly by incorporating models of preservation bias into temporal constraints (Tavaré et al., 2002). Newly developed methods exist where probability distributions can be applied to calibrated nodes in a Bayesian framework (Drummond et al., 2006; Ho, 2007; Yang and Rannala, 2006). Although we recognize that this approach is superior in that it can lend more credence to the fossil record than standard minimum-age constraints, we refrained from using such methods here as there is currently insufficient information with which to specify meaningful prior distributions for most avian diversification times. Realization of such distributions will require more communication between paleontologists and molecular phylogeneticists (Brochu et al., 2004; Donoghue and Benton, 2007).

CONCLUSION

One possible explanation for the discrepancy between molecular and fossil data in dating early divergences of avian lineages has been that the genetic data have been misinterpreted. In this vein, the ancient age estimates reported from previous molecular studies are seen as artefacts of oversimplified or improperly executed methods. Here we have examined this conjecture by accommodating uncertainty in genetic divergence time estimates. Through analyses of a large mtDNA matrix using multiple cladistically tested

calibrations, alterative tree topologies and several sophisticated relaxed clock methods we have found that the molecular estimates are robust to varying assumptions about the evolution of evolutionary rates and consistent with those from previous studies. Our findings thus strongly support pre-K-Pg genetic origins for multiple modern bird lineages, including various extant orders and families, in contrast to the model of post-K-Pg diversification derived from a narrow interpretation of the fossil record.

METHODS

Molecular sequence data

The molecular data set comprises 135 avian species and 100 traditionally recognized families, as well as five outgroup taxa (turtles, $n = 3$; crocodilians, $n = 2$; Table 3.6). The names for avian taxa shown in our figures and tables generally follow Gill and Wright (2006). Turtles were used solely to root phylogenies and were not used for dating purposes. For each taxon a total of 5248 base pairs (bp) of mitochondrial DNA (mtDNA) was either sequenced directly using primers reported by or modified from (Sorenson et al., 1999) or downloaded from GenBank. mtDNAs from taxa newly collected by us and first reported here (GenBank accession numbers EU166921-EU167086, EU372666-EU372688, EU391159). Protein-coding genes were aligned at the amino acid level, while RNA genes were aligned as nucleotides using secondary structure models following (Mindell et al., 1997). From the original matrix, 654 alignment positions could not be aligned unambiguously and so were excluded from subsequent analyses, yielding a final matrix of 4594 bp (Table 3.4).

Fossil calibration points

We include as many fossil calibration points as possible in disparate parts of the avian tree to maximize information from the fossil record, and reduce dependency on any one calibration estimate. Given that serious bias can result if even a single fossil has been misdiagnosed in its taxonomic affinity, we limit our calibration points to those fossils that have been rigorously analyzed in a cladistic context (Table 3.5). Our fossil calibrations are nearly identical to those used by Brown et al. (2007), which is a modification of the fossil complement used by Ericson et al. (2006). These internal avian calibration points

are all implemented as minimum age constraints in the various dating analyses. We also use two bounded external calibrations derived from the fossil record that date the caiman-alligator (71-66 MY) and crocodile-bird (251-243 MY) splits (Müller and Reisz, 2005). This last calibration has recently been independently supported by molecular data using soft calibration bounds (Sanders and Lee, 2007).

Phylogenetic trees and branch length uncertainty

Inferring dates on an incorrect tree topology will lead to estimates that are suspect, if not wholly incorrect. We seek to accommodate the existing uncertainty about avian phylogenetic relationships by dating nodes on two alternative trees. The first topology considered is taken from Figure 27.10 of Cracraft et al. (2004), which is a consensus tree based on previous molecular- and morphology-based phylogenetic reconstructions. This tree is conservative in that all represented branching events are well supported by independent lines of evidence; uncertain relationships among lineages are presented as hard polytomies. This topology is referred to as $T_{\text{Consensus}}$. The second topology considered was derived from a partitioned-model maximum likelihood search using the program RAxML-VI-HPC 2.2.3 (Stamatakis et al., 2005). The data were divided into four partitions: first, second and third codon positions of mitochondrial cytochrome-b, ND1 and ND2 genes, and concatenated mitochondrial 12S rDNA and tRNA genes (L, I, Q, M, W, A, N, C, Y). A collapsed consensus tree derived from Cracraft et al (thick branches only of Figure 27.10 in Cracraft et al. (2004)) was used as a backbone constraint to make tree searches more efficient ($T_{\text{Constraint}}$; see Figure 3.7). Several hundred heuristic searches were performed under the GTRMIX model assuming a range of values for both the initial rearrangement setting (*i*; range 5–20) and number of rate categories (*c*; range 5–25). The topology of the maximum likelihood estimate (MLE) is referred to as T_{Optimal}.

For the non-Bayesian dating methods, we accommodate uncertainty in branch length estimation through a non-parametric bootstrapping procedure. For each original data partition, 100 pseudoreplicate datasets were generated using the program SEQBOOT of the PHYLIP 3.6 package (Felsenstein, 2004); these bootstrapped matrices were concatenated to produce 100 pseudomatrices with the same size and number of partition-specific characters as the original matrix. For T_{Optimal}, branch lengths and substitution

model parameters were estimated using a partitioned GTR+G model in RAxML. For $T_{\text{Consensus}}$, branch lengths and all substitution model parameters were estimated from each bootstrap matrix on the fixed topology using the GTR+I+G substitution model in PAUP* (Swofford, 2003) because RAxML cannot evaluate a tree containing mutlifurcations. Using these procedures we generated 100 trees with branch lengths (phylograms) for each topology.

Divergence time estimates using relaxed molecular clocks

Owing to the concern that assumptions of particular analytical methods may systematically bias divergence time estimates, we compare several methods for accommodating among-lineage rate heterogeneity for the purpose of estimating the divergence times of modern avian lineages. To facilitate direct comparison, all methods utilize the same fossil calibrations and tree topologies outlined above.

First, the program r8s 1.7 (Sanderson, 2003) was used to estimate rates and dates for internal nodes in the phylogeny via penalized likelihood (PL). This semiparametric procedure combines a parametric model that allows for different rates on each branch in the tree (Sanderson, 2002) with a non-parametric penalty function that penalizes rates that change too quickly along the tree from ancestor to descendent branches (Sanderson, 1997); a smoothing parameter (λ) determines the relative contribution of the penalty function. The optimal smoothing value was determined individually for each topology through a sequence-based cross-validation procedure (Sanderson, 2002) using penalty = add and the normalized (χ^2) errors. Confirmation of the optimal smoothing values was obtained via multiple optimizations with different initial conditions (by setting num_restarts = num_time_guesses = 3). Confidence intervals for node ages were determined using the distribution of estimated ages across bootstrapped trees assuming the optimal smoothing value from the original phylograms. Summary of the mean estimate and associated error for a given node was accomplished using the profile command in r8s.

Second, the program PATHd8 (Britton et al., 2007; Britton et al., 2006) also smoothes rates across the tree, but does so by calculating an average path length from an

internal node to all its descending terminals; deviations from a molecular clock are corrected through reference to fossil calibrations. Important distinctions from r8s above are that PATHd8 smoothes rates between sister groups (rather than from ancestor to descendent nodes) and it does this locally rather than over the entire tree. The same 100 phylograms as analyzed with r8s above were used to generate confidence intervals on divergence times, although summary statistics were calculated manually.

As a third approach, the Bayesian MULTIDISTRIBUTE package (Thorne, 2003) represents an attempt to explicitly model rate change across a tree (Kishino et al., 2001; Thorne and Kishino, 2002; Thorne et al., 1998). Here, the logarithm of the rate at the end of a branch is modelled with a normal distribution, the mean of which has an expected value equal to the rate at the beginning of the branch; thus, rates are implicitly assumed to be autocorrelated from ancestor to descendent nodes, although this autocorrelation may decay with increasing branch lengths. The posterior probability distribution of divergence times is approximated by samples from a Markov chain Monte Carlo (MCMC) chain. The data were partitioned as noted above. For each partition, estimates of the transition/transversion rate ratio and the gamma site class-specific rates under the F84+G model (the most complex model supported by the MULTIDISTRIBUTE package) were calculated in the baseml program of the PAML 3.15 package (Yang, 1997). The output from baseml was used as the input for the MULTIDISTRIBUTE program estbranches, which produced MLEs of branch lengths and their approximate variance-covariance matrix. Although differing in branching order, $T_{\text{Consensus}}$ and T_{Optimal} had similar overall tree lengths; as a result, the same priors were applied to both topologies in the program Multidivtime: rtrate = rtratesd = 0.012, and brownmean = brownmeansd = 0.01. As one of our external calibration points bounds the root node (crocodile-bird split at 251-243 MY), date priors were less important and were defined narrowly (bigtime = 255 MY, rttm = 247 MY, rttmsd = 1.5 MY). The program was run without the assumption of correlated changes in rates across data partitions. Following a burnin of 10^6 samples, 10^4 samples were taken at a sampling interval of 10^2. All analyses were repeated with different inseed values to check for convergence of the MCMC chain.

Fourth, the overdispersed clock method of Cutler (Cutler, 2000b) represents a strong

departure from other approaches to dating in the way it models branch length heterogeneity. Instead of treating a variable number of substitutions across lineages as indicative of changes in substitution rate across the tree, Cutler's method assumes that rate is stationary, but with high variance. Thus, rather than assuming that the number of substitutions along a lineage is Poisson distributed (where the variance is equal to the mean), the observed variation across branches is accommodated by the larger variance afforded through a Gaussian distribution. As a result, branches with either particularly high or low numbers of substitutions need not be clustered on the tree; in other words, heritability (phylogeny) plays no role in the observed numbers of substitutions. The program Dating5 (Cutler, 2000a) calculates χ^2 statistics for both the constant-rate Poisson and stationary models and compares the overall fit of the models using a likelihood ratio test. In addition, the program calculates *R*, the index of dispersion (the ratio of the variance to the mean number of substitutions) under the stationary model. Dating was restricted to T_{Optimal} as the current version of Dating5 does not allow for polytomies. In addition, we could not achieve likelihood convergence with the partitioned data, and so reported results are from concatenated sequences. From asymptotic likelihood theory, 95% confidence intervals were calculated using a threshold of 2 log likelihood units around the MLE.

Finally, the program BEAST 1.4.6 (Drummond and Rambaut, 2007) differs in two important ways from the dating methods listed above. First, it does not require a fixed topology; rather, branch lengths, topology, substitution model parameters and dates can be estimated simultaneously. This incorporation of uncertainty in topology may be particularly important for the present data set, where relationships amongst many clades are uncertain. Second, BEAST does not assume that substitution rates are necessarily autocorrelated across the tree. Although intuitively satisfying, autocorrelation of rates has not been demonstrated in the literature; in fact, Drummond et al. (2006) report that many empirical data sets do *not* exhibit significant autocorrelation of rates.

BEAST 1.4.6 offers two statistical distributions for describing the change in rate across a branch (Drummond et al., 2006): rates can be drawn independently from either an exponential distribution (UCED) or a lognormal distribution (UCLN). To find which

distribution fit the present data best, we initially fixed the tree topology to $T_{\text{Consensus}}$. The data were partitioned as above, with each partition assigned a GTR+I+G substitution model. BEAST MCMC runs of 25×10^6 generations following a burnin of 10^5 generations were performed for UCED, UCLN and CLOCK models. To arbitrate between models, we calculated Bayes factors by comparing harmonic mean model likelihoods (Suchard et al., 2001). For both non-autocorrelated models, we also calculated the covariance among branch rates, which indicates the degree of ancestor-descendant autocorrelation of rates across the tree. Using the optimal model, we then accommodated topological uncertainty by removing the restriction of a fixed tree. However, we did constrain certain clades (the backbone constraints described above) to be monophyletic to facilitate the placement of calibration points and increase the efficiency of the MCMC search. Three replicate runs of 25×10^6 generations were performed to check for convergence of the MCMC chain. Mean parameter estimates and 95% highest posterior densities (HPDs) were determined through analyzing the combined BEAST tree files in TreeAnnotator 1.4.6 (Rambaut and Drummond, 2006). We refer to these results with the topology T_{Flexible}.

For each of the five dating methods above we calculated the average discrepancy between molecular (MA) and fossil (FA) estimated ages for those nodes that had fossil calibrations. The MA used in these calculations was the mean estimate from MCMC samples (Multidivtime, BEAST), or the optimal estimate from the empirical data matrix (r8s, PATHd8, Dating5). The value MA-FA is equivalent to $\delta_{Realized\ MA\text{-}FA}$ described above, and gives an indication of the degree of agreement between the molecular data and the fossil record.

Episodic evolution and information content

If present, episodic (or punctuated) molecular evolution could seriously bias molecular genetic estimates of divergence time by systematically overestimating the ages of all nodes that preceded the punctuation. We therefore tested for the presence of episodic evolution in two ways. First, we used the method of Pagel et al (Pagel et al., 2006a; Pagel et al., 2006b) which quantifies the proportional contribution of punctuated (β) and gradual (g) evolution to path lengths in a phylogeny, based on extent of association

between sequence change and cladogenesis events. This method requires a fully bifurcating tree, and so analyses were limited to our optimal reconstruction T_{Optimal}. To test for this signature we analyzed maximum likelihood trees from RAxML bootstrap analyses (n = 100). Second, to test whether substitution rate accelerated coincident with or following the K-Pg boundary we plotted standardized inferred substitution rate (per data partition) versus inferred divergence time from the Multidivtime analyses above. If the K-Pg boundary extinctions facilitated diversification of avian higher-level taxa, it could produce a spike in this plot (Aris-Brosou and Yang, 2003) detected as a departure from the mean standardized rate. These two tests are complementary in that the first focuses specifically on the effects of speciation, whereas the second focuses on absolute time effects, which may or may not be related to speciation.

As in the case of episodic evolution, a simple lack of historical signal present in molecular sequences could generate erroneous divergence time estimates. We therefore assessed the 'information content' present in our mtDNA matrix by regressing posterior 95% credible width against posterior mean divergence time. When the when the amount of data is saturated then this regression will produce a linear relationship (i.e. $R^2 \rightarrow 1$), the slope of which indicates the amount of uncertainty that cannot be reduced through adding longer sequences (Rannala and Yang, 2007; Yang and Rannala, 2006), but can be reduced through adding independent loci.

ACKNOWLEDGEMENTS

I thank A. Stamatakis (RAxML), J. Thorne (MULTIDISTRIBUTE) and D. Cutler (Dating5) for assistance with their respective software packages, and S. Ho for assistance with the new methods available in BEAST. M. van Tuinen offered indispensable advice regarding an initial set of fossil constraints and R. B. Payne provided essential comments on an earlier (encyclopaedic) draft of this manuscript. A. Baker and A. Lindsay performed some of the laboratory work. Three anonymous reviewers offered critical suggestions to improve the manuscript. JWB thanks R. Pollard for sustained encouragement throughout. Funding was provided by the National Science Foundation and the University of Michigan.

Table 3.1 Degree of autocorrelation in rates of molecular evolution by partition and tree topology as calculated in Multidivtime.

Topology[1]	Genetic partition	Autocorrelation (95% CI)[2]
$T_{\text{Consensus}}$	First positions	0.00197 (0.00127, 0.00290)
	Second positions	0.00437 (0.00258, 0.00685)
	Third positions	0.00452 (0.00288, 0.00680)
	RNA	0.00566 (0.00343, 0.00874)
T_{Optimal}	First positions	0.00177 (0.00112, 0.00263)
	Second positions	0.00344 (0.00197, 0.00548)
	Third positions	0.00380 (0.00241, 0.00571)
	RNA	0.00414 (0.00206, 0.00744)

[1] See Figure 3.2 for alternative topologies.
[2] CI = Bayesian credible interval.

Table 3.2 Model comparisons for analyses relaxing the assumption of autocorrelation of rates across the tree in BEAST. For these model comparisons, the topology was fixed as $T_{\text{Consensus}}$ (see Figure 3.2). The strict clock model serves as a base comparison. The tree T_{Flexible} refers to analyses where topology is not fixed. 95% Highest posterior densities (HPDs) are given in parentheses.

Model	Model likelihood[1]	Covariance[2]
$T_{\text{Consensus}}$		
CLOCK	-212231	N/A
UCED	-210459	0.039 (-0.103, 0.175)
UCLN	-207226	0.066 (-0.061, 0.193)
T_{Flexible}		
UCLN	-206812	0.042 (-0.071, 0.161)

[1] Harmonic mean model likelihoods were calculated from post-burnin MCMC samples.

[2] Indicates the degree of substitution rate autocorrelation between ancestor and descendent branches.

Table 3.3 Estimated divergence times for major avian clades compared across methods and topologies. Estimated time to most recent common ancestor (tMRCA) of clades of non-controversial monophyletic status. Standard deviations are given in parentheses (for Dating5 and BEAST, standard deviations were calculated from 95% confidence/credible intervals using a normal approximation). For r8s, PATHd8 and Multidivtime ages were estimated on each of the two fixed topologies ($T_{\text{Consensus}}$ and T_{Optimal}; see Figure 3.2). For BEAST, divergence times were estimated simultaneously with phylogeny (T_{Flexible}; Figure 3.4). For each method an estimate of the magnitude of disagreement between fossil and molecular estimates of divergence times ($\delta_{Realized\ MA\text{-}FA}$) is calculated as an average of MA-FA (the molecular age minus the fossil age) for those 18 internal nodes with calibration points (Table 3.5).

		r8s		**PATHd8**		**Multidivtime**		**Dating5**	**BEAST**
Node	**tMRCA**	$T_{\text{Consensus}}$	T_{Optimal}	$T_{\text{Consensus}}$	T_{Optimal}	$T_{\text{Consensus}}$	T_{Optimal}	T_{Optimal}	T_{Flexible}
A	Neognaths-Paleognaths	125.7 (12.4)	131.1 (10.7)	114.3 (6.9)	102.8 (6.1)	129.9 (11.0)	132.4 (10.7)	132.9 (11.6)	133.2 (8.1)
B	Paleognaths	98.1 (10.6)	104.8 (10.7)	72.8 (5.0)	66.3 (4.6)	107.6 (11.2)	110.1 (11.2)	80 (6.8)	105.9 (11.7)
C	Galloanserae	93.6 (10.7)	100.7 (10.1)	86.4 (5.5)	78.7 (4.7)	97.3 (9.9)	100.6 (9.5)	89.3 (3.2)	110.4 (7.8)
D	Galloanserae-Neoaves	114.6 (12.1)	121.9 (10.5)	103.1 (6.0)	93.1 (5.4)	116.1 (11.0)	120.8 (10.5)	126.8 (6.1)	126.0 (7.1)
E	Neoaves	104.5 (11.4)	116.6 (9.9)	90.4 (5.1)	86.1 (5.0)	101.3 (10.1)	113.4 (10.1)	123.9 (5.3)	118.5 (6.8)
F	Ratites	67.4 (9.6)	89.3 (12.1)	49.5 (3.5)	46.7 (3.2)	92.1 (10.3)	97.3 (10.4)	40.6 (12.3)	91.5 (12.0)
G	Galliformes	82.1 (9.7)	88.4 (9.4)	82.2 (6.0)	73.2 (5.4)	87.4 (9.5)	87.2 (9.2)	67.3 (11.3)	99.0 (8.4)
H	Anseriformes	82.7 (10.1)	89.1 (10.7)	70.6 (4.0)	67.1 (2.8)	88.5 (9.3)	91.5 (9.0)	85.4 (4.1)	100.5. (8.3)
I	Charadriiformes	81.8 (11.5)	94.0 (9.2)	55.4 (3.6)	49.9 (3.1)	80.2 (8.5)	80.6 (7.8)	41.9 (4.5)	81.7 (6.3)
J	Passeriformes	85.4 (9.1)	99.5 (8.2)	89.0 (5.5)	85.5 (5.2)	78.4 (8.5)	97.8 (9.3)	84.9 (12.5)	106.6 (7.2)
K	Piciformes	90.9 (10.1)	99.6 (9.0)	89.0 (5.5)	79.0 (4.8)	83.0 (9.2)	91.1 (9.1)	101.0 (8.8)	93.6 (6.8)
L	Procellariiformes	73.8 (10.8)	89.9 (9.1)	55.6 (2.2)	38.1 (2.8)	80.0 (8.7)	78.5 (7.8)	38.8 (12.4)	74.7 (7.3)
M	Cuculiformes	73.9 (8.3)	79.5 (7.4)	65.0 (4.4)	60.1 (4.1)	68.3 (8.0)	74.3 (8.0)	52.5 (6.7)	74.1 (8.6)
N	Strigiformes	88.2 (9.9)	94.7 (8.6)	89.0 (5.9)	79.0 (5.1)	82.5 (9.7)	88.5 (9.4)	93.2 (11.5)	84.2 (9.1)
O	Apodiformes	77.4 (9.3)	75.1 (7.4)	70.0 (5.6)	53.5 (1.9)	77.3 (9.0)	63.4 (7.2)	55.8 (9.1)	80.5 (9.9)
	$\delta_{Realized\ MA\text{-}FA}$	39.8	44.6	24.2	16.9	36.8	36.1	24.8	36.5

Table 3.4 Aligned fragment lengths of mtDNA sequences (total 4594 bp). Codon positions are combined across all protein-coding genes (ND1, ND2, cytochrome b), and RNA includes 12S and nine tRNA genes (L, I, Q, M, W, A, N, C, Y).

Gene	Aligned length (bp)	Variable sites (%)	Parsimony informative sites (%)
1st codon positions	1043	699 (67%)	605 (58%)
2nd codon positions	1043	512 (49%)	380 (36%)
3rd codon positions	1043	1043 (100%)	1041 (100%)
RNA	1465	910 (62%)	728 (50%)

Table 3.5 Fossil calibrations. All internal calibrations for Neornithes are treated as minimum ages. External calibrations are treated as bounded (lower, upper) constraints. See Ericson et al. (2006) for fossil citations and details. See Figure 3.2 for placement of calibrations in the alternative trees.

Fossil ID	Calibration	Age (MY)	Source
1	Crown Pici	30	Ericson et al. (2006)
2	Stem Upupidae + Phoeniculidae	47.5	Ericson et al. (2006)
3	Stem Coraciidae + Brachypteraciidae	47.5	Ericson et al. (2006)
4	Stem Trogoniformes	53	Ericson et al. (2006)
5	Stem Coliiformes	55	Ericson et al. (2006)
6	Stem Strigiformes	55	Ericson et al. (2006)
7	Crown Pandionidae	37	Ericson et al. (2006)
8	Stem Anatidae	66	Clarke et al. (2005)
9	Crown Sulidae	33	Ericson et al. (2006)
10	Stem Fregatidae	53	Ericson et al. (2006)
11	Stem Sphenisciformes	62	Slack et al. (2006)
12	Stem Jacanidae	30	Ericson et al. (2006)
13	Stem Apodiformes	53	Ericson et al. (2006)
14	Stem Trochilidae	30	Mayr (2004)
15	Crown Pteroclididae	30	Ericson et al. (2006)
16	Stem Phoenicopteriformes	30	Ericson et al. (2006)
17	Stem Phaethontidae	55	Ericson et al. (2006)
18	Stem Gruidae	30	Ericson et al. (2006)
20	Alligator-caiman	66–71	Müller and Reisz (2005)
21	Bird-crocodile	243–251	Müller and Reisz (2005)

Table 3.6 Sample information for taxa used in this study.

Order	Family	Species	Common Name	12S	CYTB	ND1	ND2	tRNAs
Testudines								
	Cheloniidae	*Chelonia mydas*	Green sea-turtle	AB012104	AB012104	AB012104	AB012104	AB012104
	Emydidae	*Chrysemys picta*	Painted turtle	NAF069423	NAF069423	NAF069423	NAF069423	NAF069423
	Pelomedusidae	*Pelomedusa subrufa*	African helmeted turtle	AF039066	AF039066	AF039066	AF039066	AF039066
Crocodylia								
	Crocodylidae	*Alligator mississipiensis*	American alligator	Y13113	Y13113	Y13113	Y13113	Y13113
	Crocodylidae	*Caiman crocodilus*	Spectacled caiman	AJ404872	AJ404872	AJ404872	AJ404872	AJ404872
			Paleognathae					
Struthioniformes	Apterygidae	*Apteryx haastii*	Great spotted kiwi	AF338708	AF338708	AF338708	AF338708	AF338708
Struthioniformes	Casuariidae	*Casuarius casuarius*	Southern cassowary	AF338713	AF338713	AF338713	AF338713	AF338713
Struthioniformes	Dromaiidae	*Dromaius novaehollandiae*	Emu	AF338711	AF338711	AF338711	AF338711	AF338711
Struthioniformes	Dinornithidae	*Dinornis giganteus*	Giant moa	AY016013	AY016013	AY016013	AY016013	AY016013
Struthioniformes	Emeidae	*Anomalopteryx didiformis*	Little bush moa	AF338714	AF338714	AF338714	AF338714	AF338714
Struthioniformes	Emeidae	*Emeus crassus*	Eastern moa	AY016015	AY016015	AY016015	AY016015	AY016015
Struthioniformes	Rheidae	*Pterocnemia pennata*	Lesser rhea	AF338709	AF338709	AF338709	AF338709	AF338709
Struthioniformes	Rheidae	*Rhea americana*	Greater rhea	AF090339	AF090339	AF090339	AF090339	AF090339
Struthioniformes	Struthionidae	*Struthio camellus*	Ostrich	AH007281	AH007281, Y12025	AH007281	AH007281	AF069429
Tinamiformes	Tinamidae	*Crypturellus undulatus*	Undulated tinamou	AY139627	AY139629	AY139628	AY139628	AY139628
Tinamiformes	Tinamidae	*Eudromia elegans*	Elegant crested tinamou	AY274002	AY274002	AY274002	AY274002	AY274049
Tinamiformes	Tinamidae	*Tinamus major*	Great tinamou	AF338707	AF338707	AF338707	AF338707	AF338707
			Galloanserae					
Anseriformes	Anatidae	*Anser albifrons*	Greater white-fronted goose	AF363031	AF363031	AF363031	AF363031	AF363031
Anseriformes	Anatidae	*Aythya americana*	Redhead	AF069422	AF069422	AF069422	AF069422	AF090337
Anseriformes	Anatidae	*Branta canadensis*	Canada goose	DQ019124	DQ019124	DQ019124	DQ019124	DQ019124
Anseriformes	Anhimidae	*Chauna torquata*	Southern screamer	AY274030	AY274006	AY274053	AY274053	AY274053
Anseriformes	Anseranatidae	*Anseranas semipalmata*	Magpie goose	AF173772	U83730	AY274054	AY274054	AY274054
Anseriformes	Dendrocygnidae	*Dendrocygna arcuata*	Wandering whistling-duck	AF536743	AF536740	AF536746	AF536746	U97735
Galliformes	Cracidae	*Crax rubra*	Great curassow	AY274029	AY274003	AY274050	AY274050	AY274050
Galliformes	Megapodidae	*Megapodius eremita*	Melanesian megapode	AF082065	AY274005	AY274052	AY274052	AY274052
Galliformes	Megapodidae	*Alectura lathami*	Brush turkey	AF082058	AY274004	AY274051	AY274051	AY346091

Galliformes	Numididae	*Acryllium vulturinum*	Vulturine guineafowl	AF536742	AF536739	AF536745	AF536745	AF536745
Galliformes	Numididae	*Numida meleagris*	Helmeted guineafowl	AP005595	AP005595	AP005595	AP005595	AP005595
Galliformes	Odontophoridae	*Colinus virginianus*	Northern bobwhite	EU167061	EU372675	EU166949	EU166949	EU166949
Galliformes	Phasianidae	*Coturnix chinensis*	King quail	AB073301	AB073301	AB073301	AB073301	AB073301
Galliformes	Phasianidae	*Coturnix japonica*	Japanese quail	AP003195	AP003195	AP003195	AP003195	AP003195
Galliformes	Phasianidae	*Gallus gallus*	Red junglefowl	X52392	X52392	X52392	X52392	X52392
			Neoaves					
Apodiformes	Apodidae	*Aeronautes saxatalis*	White-throated swift	EU167032	EU166978	EU166921	EU166921	EU166921
Apodiformes	Trochilidae	*Metallura eupogon*	Fire-throated metaltail	EU167083	EU167027	EU166922	EU166922	EU166922
Apodiformes	Trochilidae	*Phaethornis syrmathophorus*	Tawny-bellied hermit	EU167084	EU167028	EU166923	EU166923	EU166923
Caprimulgiformes	Caprimulgidae	*Chordeiles minor*	Common nighthawk	EU167037	EU166983	EU166924	EU166924	EU166924
Caprimulgiformes	Eurystopodidae	*Eurostopodus macrotis*	Great eared-nightjar	EU167043	EU166989	EU166925	EU166925	EU166925
Caprimulgiformes	Nyctibidae	*Nyctibius maculosus*	Andean potoo	EU167060	EU167006	EU166926	EU166926	EU166926
Caprimulgiformes	Podargidae	*Podargus strigoides*	Tawny frogmouth	EU167069	EU167014	EU166927	EU166927	EU166927
Caprimulgiformes	Steatornithidae	*Steatornis caripensis*	Oilbird	EU167079	EU167023	EU166928	EU166928	EU166928
Charadriiformes	Alcidae	*Cepphus columba*	Pigeon guillemot	EU372666	EU372673	EU372680	EU372680	EU372680
Charadriiformes	Burhinidae	*Burhinus senegalensis*	Senegal thick-knee	AY274043	AY274007	AY274073	AY274073	AY274073
Charadriiformes	Charadriidae	*Charadrius semipalmatus*	Semipalmated plover	EU167040	EU166986	EU166929	EU166929	EU166929
Charadriiformes	Glareolidae	*Glareola pratincola*	Collared pratincole	EU372667	EU372674	EU372681	EU372681	EU372681
Charadriiformes	Haematopodidae	*Haematopus ater*	Blackish oystercatcher	AY074886	AY074886	AY074886	AY074886	AY074886
Charadriiformes	Haematopodidae	*Haematopus ostralegus*	Eurasian oystercatcher	EU167052	EU166998	EU166930	EU166930	EU166930
Charadriiformes	Jacanidae	*Jacana jacana*	Wattled jacana	EU167053	EU166999	EU166935	EU166935	EU166935
Charadriiformes	Laridae	*Larus atricilla*	Laughing gull	EU167055	EU167001	EU166931	EU166931	EU166931
Charadriiformes	Laridae	*Larus dominicanus*	Kelp gull	AY293619	AY293619	AY293619	AY293619	AY293619
Charadriiformes	Recurvirostridae	*Himantopus mexicanus*	Black-necked stilt	EU167077	EU167022	EU166932	EU166932	EU166932
Charadriiformes	Scolopacidae	*Arenaria interpres*	Ruddy turnstone	AY074885	AY074885	AY074885	AY074885	AY074885
Charadriiformes	Scolopacidae	*Scolopax minor*	American woodcock	AF082068	U83744	AY274072	AY274072	AY274072
Charadriiformes	Stercorariidae	*Stercorarius skua*	Pomarine skua	EU167080	EU167024	EU166933	EU166933	EU166933
Charadriiformes	Thinocoridae	*Attagis gayi*	Rufous-bellied seedsnipe	EU167081	EU167025	EU166934	EU166934	EU166934
Ciconiiformes	Ardeidae	*Nyctanassa violacea*	Yellow-crowned night heron	EU167033	EU166979	EU166936	EU166936	EU166936
Ciconiiformes	Ardeidae	*Tigrisoma fasciatum*	Fasciated tiger heron	EU167034	EU166980	EU166937	EU166937	EU166937
Ciconiiformes	Ciconiidae	*Ciconia boyciana*	Oriental wood stork	AB026193	AB026193	AB026193	AB026193	AB026193
Ciconiiformes	Ciconiidae	*Ciconia ciconia*	European wood stork	AB026818	AB026818	AB026818	AB026818	AB026818

Ciconiiformes	Ciconiidae	*Mycteria americana*	Wood stork	AF082066	U83712	AY274076	AY274076	AY274076
Ciconiiformes	Scopidae	*Scopus umbretta*	Hamerkop	EU372669	U08936	EU372682	EU372682	EU372682
Ciconiiformes	Threskiornithidae	*Eudocimus albus*	White ibis	EU167082	EU167026	EU166938	EU166938	EU166938
Coliiformes	Coliidae	*Colius striatus*	Spectacled mousebird	AY274032	AY274011	AY274058	AY274058	AY274058
Coliiformes	Coliidae	*Urocolius macrourus*	Blue-naped mousebird	AY274033	AY274012	AY274059	AY274059	AY274059
Columbiformes	Columbidae	*Columba leucocephala*	White-crowned pigeon	AY274041	AY274023	AY274070	AY274070	AY274070
Columbiformes	Columbidae	*Treron sieboldii*	White-bellied green-pigeon	AY274042	AY274024	AY274071	AY274071	AY274071
Columbiformes	Pteroclidae	*Pterocles coronatus*	Crowned sandgrouse	EU167073	EU167018	EU166939	EU166939	EU166939
Coraciiformes	Alcedinidae	*Ispidina picta*	African kingfisher	EU167031	EU166977	EU166940	EU166940	EU166940
Coraciiformes	Bucerotidae	*Tockus erythrorhynchus*	Red-billed hornbill	AF082071	AY274008	AY274055	AY274055	AY274055
Coraciiformes	Cerylidae	*Megaceryle alcyon*	Belted kingfisher	EU167039	EU166985	EU166945	EU166945	EU166945
Coraciiformes	Coraciidae	*Coracias spatulata*	Racquet-tailed roller	AF082060	AY274010	AY274057	AY274057	AY274057
Coraciiformes	Meropidae	*Merops viridis*	Blue-throated bee-eater	EU167057	EU167003	EU166941	EU166941	EU166941
Coraciiformes	Momotidae	*Momotus momota*	Blue-crowned motmot	EU167058	EU167004	EU166942	EU166942	EU166942
Coraciiformes	Phoeniculidae	*Phoeniculus purpureus*	Green wood hoopoe	EU167067	EU167012	EU166943	EU166943	EU166943
Coraciiformes	Upupidae	*Upupa epops*	Eurasian hoopoe	EU167086	EU167030	EU166944	EU166944	EU166944
Cuculiformes	Coccyzidae	*Coccyzus erythropthalmus*	Black-billed cuckoo	AF082048	AY274015	AY274062	AY274062	AY274062
Cuculiformes	Crotophagidae	*Crotophaga ani*	Smooth-billed ani	AY274035	AY274016	AY274063	AY274063	AY274063
Cuculiformes	Cuculidae	*Cuculus canorus*	Common cuckoo	AY274034	AY274013	AY274060	AY274060	AY274060
Cuculiformes	Neomorphidae	*Neomorphus geoffroyi*	Rufous-vented ground cuckoo	AY274036	AY274017	AY274064	AY274064	AY274064
Falconiformes	Accipitridae	*Buteo jamaicensis*	Red-tailed hawk	AY274044	U83720	AY274074	AY274074	AY274074
Falconiformes	Cathartidae	*Cathartes aura*	Turkey vulture	EU167038	EU166984	EU166946	EU166946	EU166946
Falconiformes	Falconidae	*Falco peregrinus*	Peregrine falcon	AF090338	AF090338	AF090338	AF090338	AF090338
Falconiformes	Pandionidae	*Pandion haliaetus*	Osprey	EU167063	EU167008	EU166947	EU166947	EU166947
Falconiformes	Sagittaridae	*Sagittarius serpentarius*	Secretary bird	EU167078	AJ604483	EU166948	EU166948	EU166948
Gaviiformes	Gaviidae	*Gavia adamsi*	Yellow-billed loon	EU167048	EU166994	EU166951	EU166951	EU166951
Gaviiformes	Gaviidae	*Gavia arctica*	Arctic loon	AY139633	AY139635	AY139634	AY139634	AY139634
Gaviiformes	Gaviidae	*Gavia immer*	Common loon	EU167047	EU166993	EU166950	EU166950	EU166950
Gaviiformes	Gaviidae	*Gavia pacifica*	Pacific loon	EU167049	EU166995	EU166952	EU166952	EU166952
Gaviiformes	Gaviidae	*Gavia stellata*	Red-throated loon	EU167050	EU166996	EU166953	EU166953	EU166953
Gruiformes	Gruidae	*Grus canadensis*	Sandhill crane	EU167051	EU166997	EU166954	EU166954	EU166954
Gruiformes	Otididae	*Eupodotis senegalensis*	White-bellied bustard	EU167062	EU167007	EU166955	EU166955	EU166955
Gruiformes	Rallidae	*Fulica americana*	American coot	EU167074	EU167019	EU166956	EU166956	EU166956

Musophagiformes	Musophagidae	*Crinifer piscator*	Western grey plantain-eater	AY274040	AY274021	AY274068	AY274068	AY274068
Musophagiformes	Musophagidae	*Musophaga violacea*	Violet turaco	AY274039	AY274020	AY274067	AY274067	AY274067
Opisthocomiformes	Opisthocomidae	*Opisthocomus hoazin*	Hoatzin	AY274027	AY274048	AF076363	AF076363	AF076363
Passeriformes	Acanthisittidae	*Acanthisitta chloris*	Rifleman	AY325307	AY325307	AY325307	AY325307	AY325307
Passeriformes	Corvidae	*Corvus frugilegus*	Rook	Y18522	Y18522	Y18522	Y18522	Y18522
Passeriformes	Dendrocolaptidae	*Lepidocolaptes wagleri*	Scaled woodcreeper	EU167041	EU166987	EU166957	EU166957	EU166957
Passeriformes	Eurylaimidae	*Smithornis sharpei*	Grey-headed broadbill	AF090340	AF090340	AF090340	AF090340	AF090340
Passeriformes	Formicariidae	*Grallaria squamigera*	Undulated antpitta	AY139636	AY139638	AY139637	AY139637	AY139637
Passeriformes	Furnariidae	*Cranioleuca baroni*	Baron's spinetail	EU167045	EU166991	EU166958	EU166958	EU166958
Passeriformes	Laniidae	*Lanius collurio*	Red-backed shrike	EU167054	EU167000	EU166959	EU166959	EU166959
Passeriformes	Motacillidae	*Motacilla alba*	White wagtail	EU167059	EU167005	EU166960	EU166960	EU166960
Passeriformes	Paridae	*Parus major*	Great tit	EU167064	EU167009	EU166961	EU166961	EU166961
Passeriformes	Picnonotidae	*Hypsipetes amaurotis*	Brown-eared bulbul	EU167068	EU167013	EU166962	EU166962	EU166962
Passeriformes	Sylviidae	*Phylloscopus occipitalis*	Western crowned-warbler	EU372671	EU372678	EU372683	EU372683	EU372683
Passeriformes	Thraupidae	*Hemispingus frontalis*	Oleaginous hemispingus	AY139639	EU372679	AY139640	AY139640	AY139640
Passeriformes	Turdidae	*Catharus guttatus*	Hermit thrush	EU372672	X74261	EU372684	EU372684	EU372684
Passeriformes	Tyrannidae	*Sayornis phoebe*	Eastern phoebe	AF536744	AF536741	AF536747	AF536747	AF536747
Passeriformes	Viduidae	*Vidua chalybeata*	Village indigobird	AF090341	AF090341	AF090341	AF090341	AF090341
Passeriformes	Zosteropidae	*Zosterops japonica*	Japanese white-eye	AY136569	EU391159	EU372685	AY136599	EU372685
Pelecaniformes	Fregatidae	*Fregata aquila*	Ascension frigatebird	EU167044	EU166990	EU166963	EU166963	EU166963
Pelecaniformes	Phaethontidae	*Phaethon rubricauda*	Red-tailed tropicbird	EU167065	EU167010	EU166964	EU166964	EU166964
Pelecaniformes	Phalacrocoracidae	*Phalacrocorax pelagicus*	Pelagic cormorant	EU167066	EU167011	EU166965	EU166965	EU166965
Pelecaniformes	Sulidae	*Sula dactylatra*	Masked booby	EU372670	EU372677	EU372686	EU372686	EU372686
Phoenicopteriformes	Phoenicopteridae	*Phoenicopterus ruber*	Greater flamingo	AY274045	U83714	AY274075	AY274075	AY274075
Piciformes	Bucconidae	*Malacoptila semicincta*	Semicollared puffbird	EU167035	EU166981	EU166966	EU166966	EU166966
Piciformes	Capitonidae	*Trachyphonus usambiro*	Usambiro barbet	EU167036	EU166982	EU166967	EU166967	EU166967
Piciformes	Galbulidae	*Galbula pastazae*	Coppery-chested jacamar	EU167046	EU166992	EU166968	EU166968	EU166968
Piciformes	Lybiidae	*Pogoniulus pusilus*	Red-fronted tinkerbird	EU167056	EU167002	EU166971	EU166971	EU166971
Piciformes	Picidae	*Colaptes auratus*	Northern flicker	EU372668	EU372676	EU372687	EU372687	EU372687
Piciformes	Ramphastidae	*Andigena cuculata*	Hooded mountain-toucan	EU167076	EU167021	EU166970	EU166970	EU166970
Piciformes	Ramphastidae	*Selenidera reinwardtii*	Golden-collared toucanet	EU167075	EU167020	EU166969	EU166969	EU166969
Podicipediformes	Podicipedidae	*Podiceps caspicus*	Eared grebe	EU167071	EU167016	EU166973	EU166973	EU166973
Podicipediformes	Podicipedidae	*Podilymbus podiceps*	Pied-billed grebe	EU167070	EU167070	EU166972	EU166972	EU166972

Procellariiformes	Diomedeidae	*Diomedea melanophrys*	Black-browed albatross	AY158677	AY158677	AY158677	AY158677	AY158677
Procellariiformes	Diomedeidae	*Diomedea nigripes*	Black-footed albatross	EU167042	EU166988	EU166974	EU166974	EU166974
Procellariiformes	Procellariidae	*Calonectris diomedea*	Cory's shearwater	AY139624	AY139626	AY139625	AY139625	AY139625
Procellariiformes	Procellariidae	*Pterodroma brevirostris*	Kerguelen petrel	AY158678	AY158678	AY158678	AY158678	AY158678
Procellariiformes	Procellariidae	*Pterodroma hasitata*	Black-capped petrel	EU167072	EU167017	EU166975	EU166975	EU166975
Psittaciformes	Psittacidae	*Nandayus nenday*	Nanday parakeet	AY274038	AY274019	AY274066	AY274066	AY274066
Psittaciformes	Psittacidae	*Neophema elegans*	Elegant parrot	AY274037	AY274018	AY274065	AY274065	AY274065
Psittaciformes	Psittacidae	*Strigops habroptilus*	Kakapo	AY309456	AY309456	AY309456	AY309456	AY309456
Sphenisciformes	Spheniscidae	*Aptenodytes patagonicus*	King penguin	AY139621	AY139623	AY139622	AY139622	AY139622
Sphenisciformes	Spheniscidae	*Eudyptes.chrysocome*	Rockhopper penguin	AY139630	AY139632	AY139631	AY139631	AY139631
Sphenisciformes	Spheniscidae	*Eudyptula minor*	Little penguin	AF362763	AF362763	AF362763	AF362763	AF362763
Strigiformes	Strigidae	*Asio otus*	Long-eared owl	AY274022	AF08206	AY274069	AY274069	AY274069
Strigiformes	Strigidae	*Ninox novaeseelandiae*	Morepork	AY309457	AY309457	AY309457	AY309457	AY309457
Strigiformes	Tytonidae	*Tyto alba*	Barn owl	EU167085	EU167029	EU166976	EU166976	EU166976
Trogoniformes	Trogonidae	*Harpactes ardens*	Philippine trogon	U94810	U94796	EU372688	EU372688	EU372688
Trogoniformes	Trogonidae	*Trogon curucui*	Blue-crowned trogon	AY274031	AY274009	AY274056	AY274056	AY274056

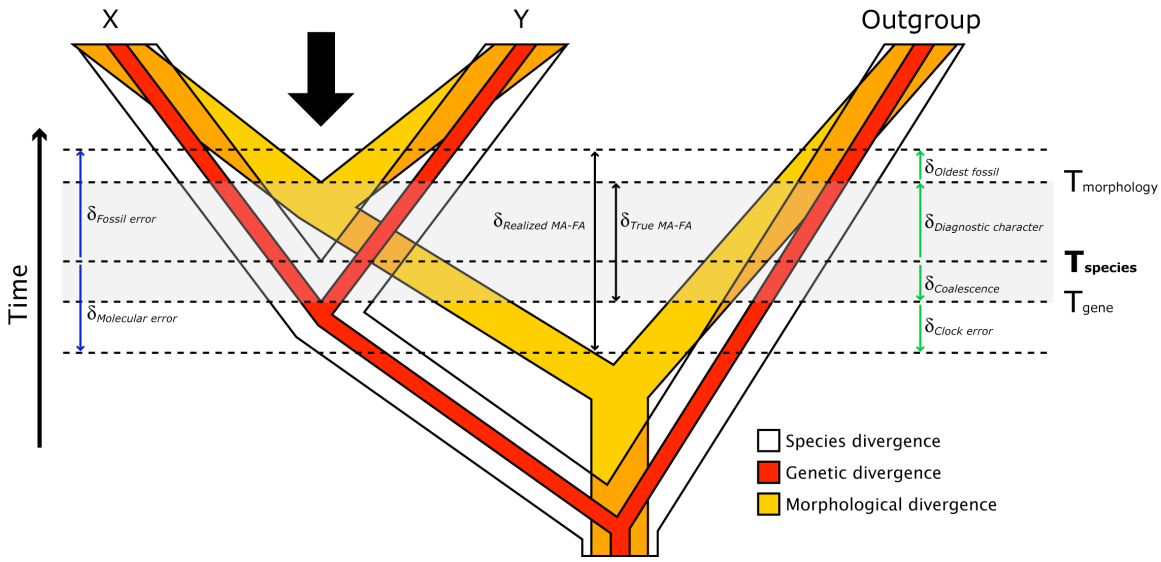

Figure 3.1 Different ways that fossil and molecular data date lineages. Time intervals defined by the horizontal dashed lines and vertical arrows pertain to age estimates for the divergence between hypothetical lineages X and Y. Even with a complete fossil record and perfect molecular clock a discrepancy is expected between fossil (FA) and molecular (MA) age estimates. As diagnostic morphological characters generally evolve ($T_{\text{Morphology}}$) after species divergence (T_{Species}), the fossil record will always underestimate (by $\delta_{Diagnostic\ character}$) the true speciation time. Genetic data, on the other hand, will overestimate speciation time (by $\delta_{Coalescence}$), as polymorphisms present during species divergence will coalesce some time in the past (T_{Gene}; related to the ancestral species effective population size). The genuine difference between molecular and morphological divergence times will thus be $\delta_{True\ MA\text{-}FA}$. With a less complete fossil record, the oldest known fossil is unlikely to temporally correspond precisely to the origination of a diagnostic character delimiting X and Y, further decreasing FA by $\delta_{Oldest\ fossil}$. Under the more realistic scenario of lineage-specific rate heterogeneity and limited taxon/character sampling, errors associated with molecular methods ($\delta_{Clock\ error}$) may result in overestimation or underestimation of the true speciation time, although underestimates are bounded by the fossil constraint ($\delta_{Fossil\ error}$). The observed discrepancy in age estimates, $\delta_{Realized\ MA\text{-}FA}$, may be considerably larger than expectations ($\delta_{True\ MA\text{-}FA}$).

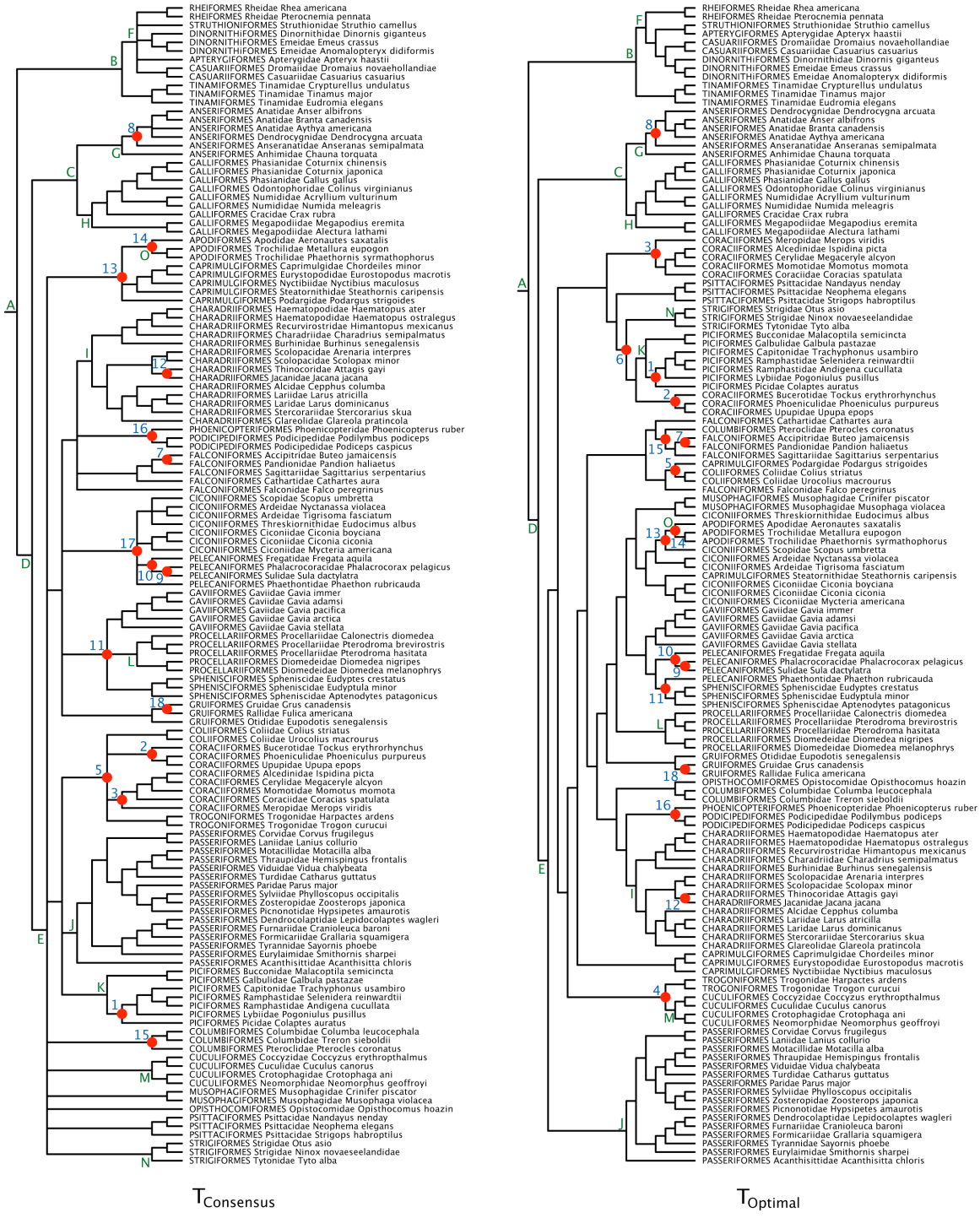

Figure 3.2 Alternative tree topologies. $T_{\text{Consensus}}$ (left) represents a conservative consensus estimate of avian familial relationships (Cracraft et al., 2004) ($AIC_c = 421460.9166$). T_{Optimal} (right) is our optimal topology derived from a partitioned model maximum likelihood search in RAxML ($AIC_c = 414160.2536$). Some topological constraints were implemented in this search (see Figure 3.7). Solid circles and numbers indicate the placement of calibration points (see Table 3.5 for ages). Letters denote nodes whose age estimates are provided in Table 3.3.

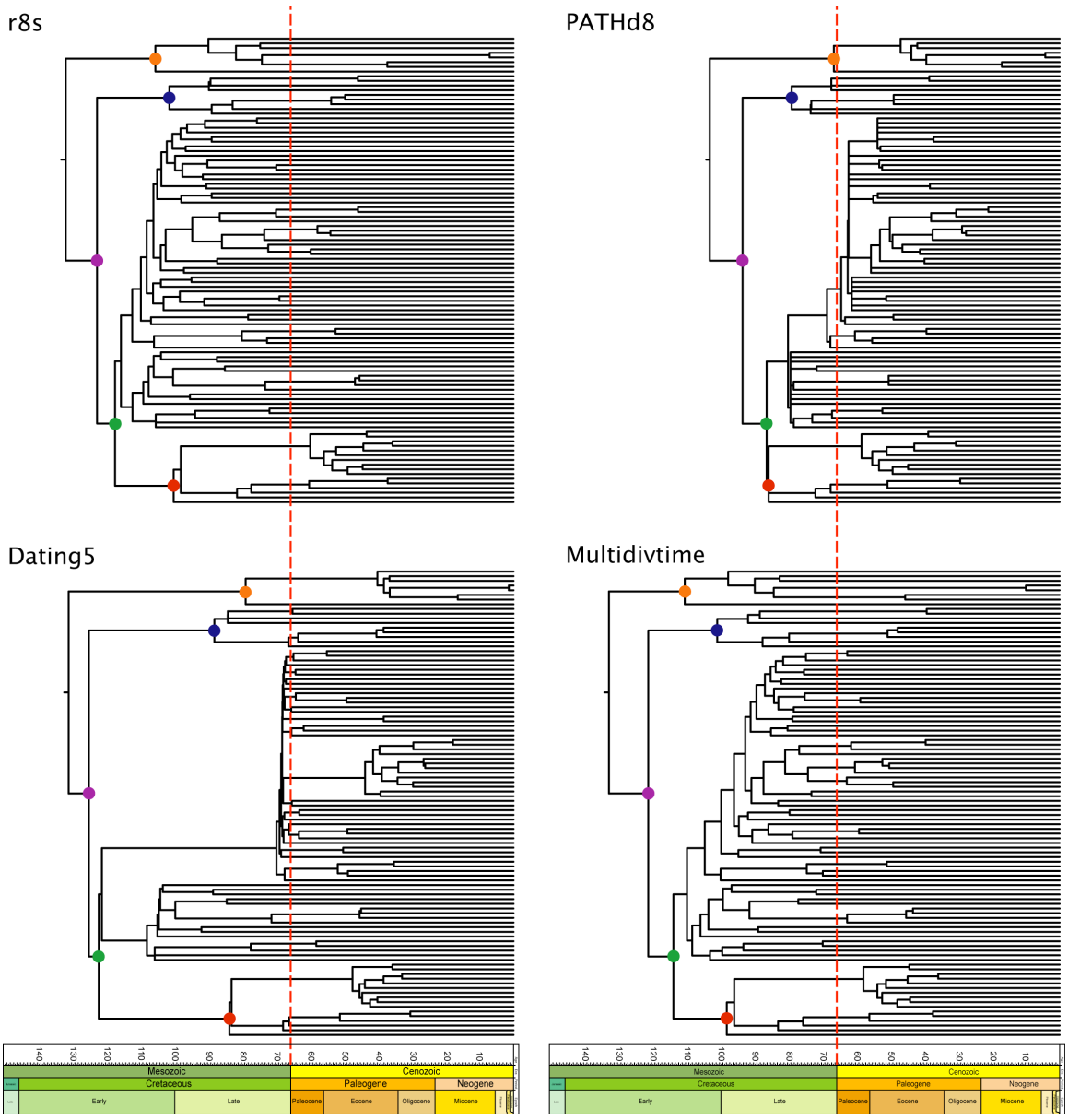

Figure 3.3 Comparative timing of divergences for avian orders and families based on four different 'relaxed clock' methods. Chronograms based on the optimal mtDNA tree reconstruction (T_{Optimal}) using r8s (top left), Dating5 (bottom left), PATHd8 (top right) and Multidivtime (bottom right); see methods for explanation of differences between analytical approaches. For legibility, error bars are removed and trees are pruned to the family level. Filled circles denote major clades: orange, Paleognathae; purple, Neognathae; blue, Galloanserae; green, Neoaves; red, Passeriformes. Time is given in millions of years before present. The vertical dashed lines indicate the K-Pg boundary. r8s, Dating5 and Multidivtime reconstructions support Cretaceous origin and diversification. PATHd8 alone supports Cretaceous origin but Tertiary diversification.

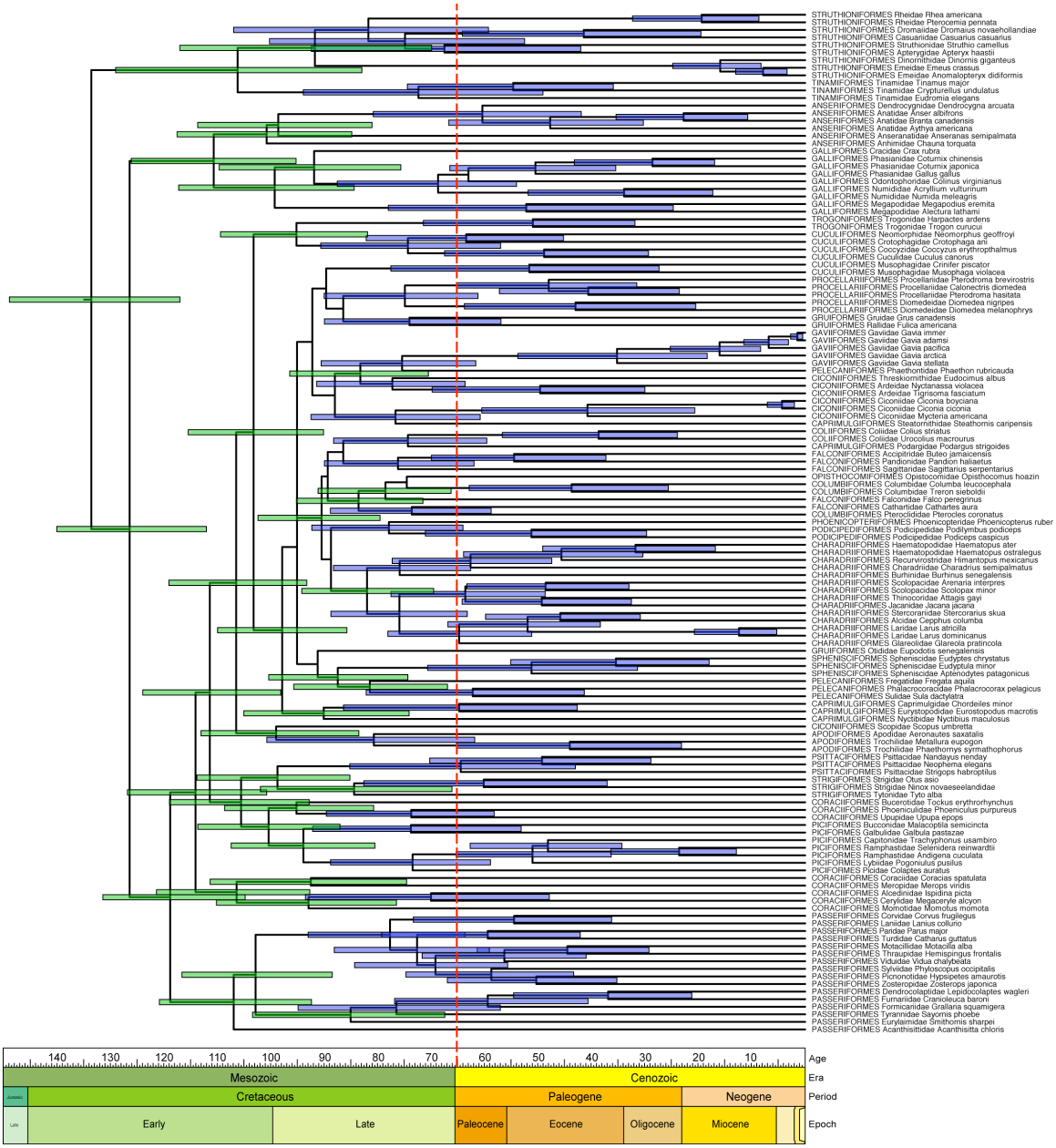

Figure 3.4 A timeline for early avian evolution. Maximum clade credible (MCC) chronogram inferred using the non-autocorrelated model of rate evolution in BEAST while allowing topology to vary (T_{Flexible}). Time is given in millions of years before present. The vertical dashed line indicates the K-Pg boundary. Error bars (blue and green) represent 95% posterior credible intervals and are only given for nodes that were present on more than 50% of the posterior sampled trees. An unambiguous ancient diversification is indicated by 37 credible intervals restricted to the Cretaceous (green bars).

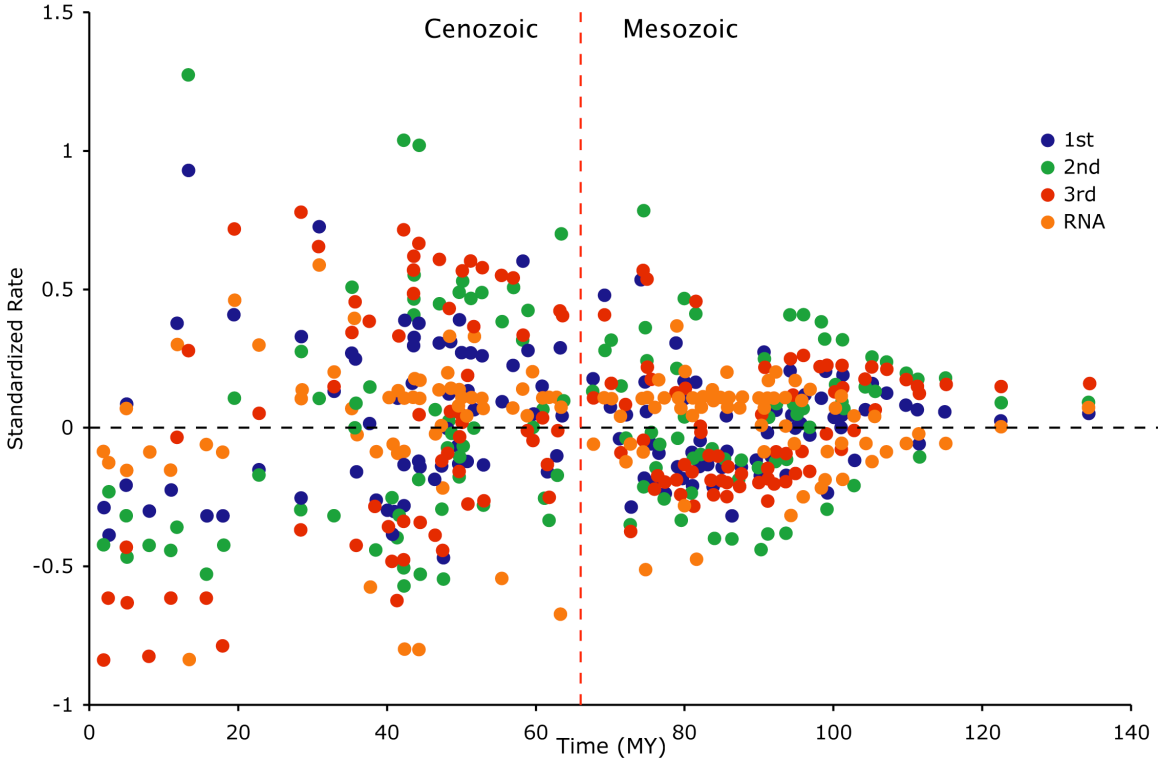

Figure 3.5 Estimated rates of molecular evolution over time, in assessment of possible episodic evolution. Standardized inferred rate of sequence evolution (per data partition) is plotted against inferred age for internal nodes on the optimal mtDNA tree reconstruction (T_{Optimal}) using Multidivtime. Time is given in millions of years before present. No support is shown for an accelerated rate accompanying initial avian diversification or following the K-Pg boundary (vertical dashed line).

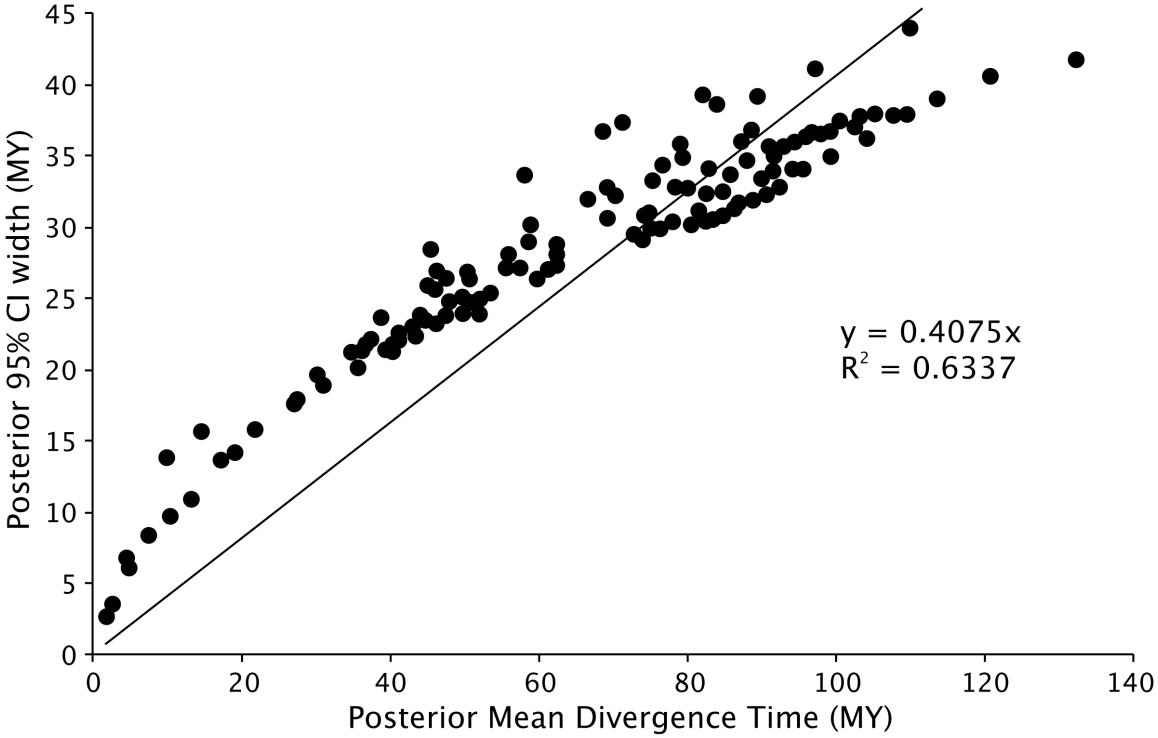

Figure 3.6 Information content of mtDNA sequences. Posterior 95% credible interval width is plotted against posterior mean divergence time using the results from Multidivtime on T_{Optimal}. Here R^2 indicates the amount of information present in the data matrix and the regression coefficient is an estimate of the expected uncertainty that is independent of sequence length.

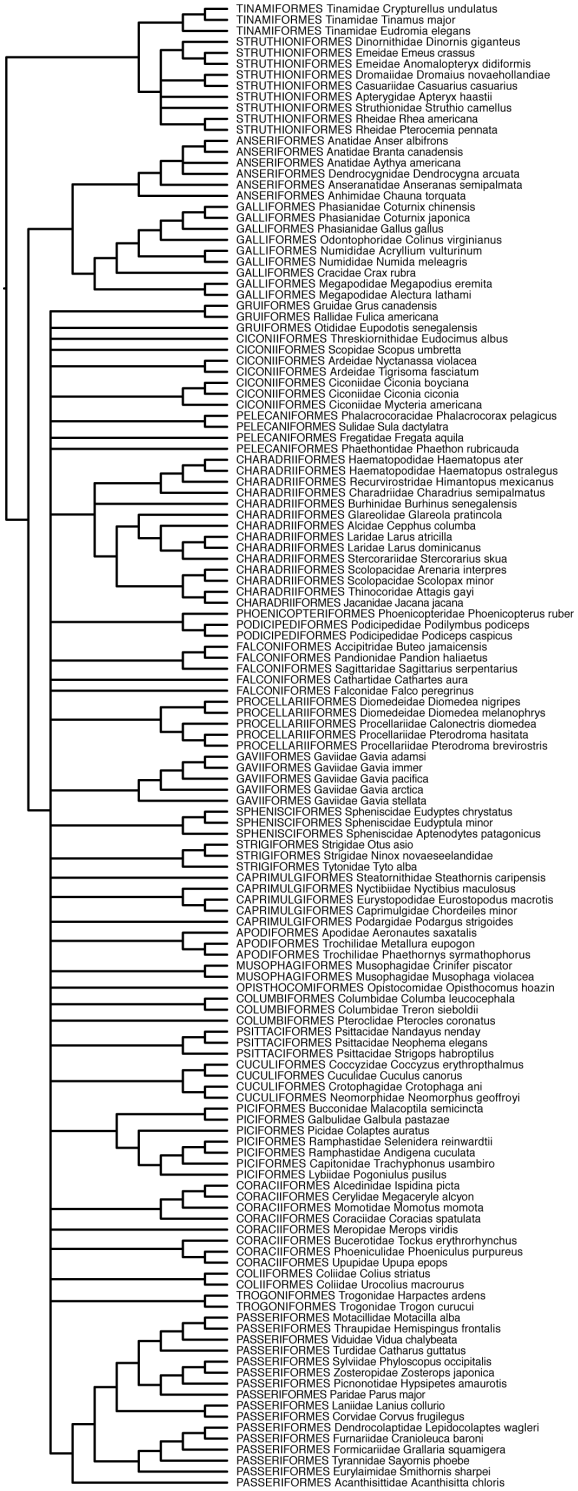

Figure 3.7 Backbone constraint topology for RAxML maximum likelihood tree searches. Derived from consensus tree of Cracraft et al. (2004; their Figure 27.10, thick branches only).

REFERENCES

Archibald, J. D., and D. H. Deutschman. 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., and Z. Yang. 2003. Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. Molecular Biology and Evolution 20:1947-1954.

Baker, A. J., S. L. Pereira, and T. A. Paton. 2007. Phylogenetic relationships and divergence times of Charadriiformes genera: multigene evidence for the Cretaceous origin of at least 14 clades of shorebirds. Biology Letters 3:205-209.

Barraclough, T. G., and V. Savolainen. 2001. Evolutionary rates and species diversity in flowering plants. Evolution 55:677-683.

Barrowclough, G. F., J. G. Groth, and L. A. Mertz. 2006. The RAG-1 exon in the avian order Caprimulgiformes: Phylogeny, heterozygosity, and base composition. Molecular Phylogenetics and Evolution 41:238-248.

Benton, M. J. 1997. Vertebrate palaeontology, 2nd edition. Chapman & Hall, London.

Benton, M. J. 1999. Early origins of modern birds and mammals: molecules vs. morphology. BioEssays 21:1043-1051.

Benton, M. J., and F. J. Ayala. 2003. Dating the tree of life. Science 300:1698-1700.

Benton, M. J., M. A. Wills, and R. Hitchin. 2000. Quality of the fossil record through time. Nature 403:534-537.

Bininda-Emonds, O. R. P., M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. Nature 446:507-512.

Bleiweiss, R. 1998. Fossil gap analysis supports early Tertiary origin of trophically diverse avian orders. Geology 26:323-326.

Britton, T., C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer. 2007. Estimating divergence times in large phylogenetic trees. Systematic Biology 56:741–752.

Britton, T., C. L. Anderson, D. Jaquet, S. Lundqvist, and K. Bremer. 2006. PATHd8—a new method for estimating divergence times in large phylogenetic trees without a molecular clock. Available from the authors (www.math.su.se/PATHd8).

Brochu, C. A., C. D. Sumrall, and J. M. Theodor. 2004. When clocks (and communities) collide: estimating divergence time from molecules and the fossil record. Journal of Paleontology 78:1-6.

Bromham, L., and D. Penny. 2003. The modern molecular clock. Nature Reviews Genetics 4:216-224.

Bromham, L., and M. Woolfit. 2004. Explosive radiations and the reliability of molecular clocks: island endemic radiations as a test case. Systematic Biology 53:758-766.

Brown, J. W., R. B. Payne, and D. P. Mindell. 2007. Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson et al. Biology Letters 3:257-259.

Chatterjee, S. 2002. The morphology and systematics of Polarornis, a Cretaceous loon (Aves: Gaviidae) from Antarctica. Pages 125-155 *in* Proceedings of the 5th Symposium of the Society of Avian Paleontology and Evolution, Beijing, 1–4 June 2000 (Z. Zhou, and F. Zhang, eds.). Science Press, Beijing.

Chiappe, L. M., and G. J. Dyke. 2002. The Mesozoic radiation of birds. Annual Review of Ecology and Systematics 33:91-124.

Clarke, J. A., and L. M. Chiappe. 2001. A new carinate bird from the late Cretaceous of Patagonia. American Museum Novitates 3323:1-23.

Clarke, J. A., C. P. Tambussi, J. I. Noriega, G. M. Erikson, and R. A. Ketcham. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. Nature 433:305-308.

Cooper, A., and R. Fortey. 1998. Evolutionary explosions and the phylogenetic fuse. Trends in Ecology & Evolution 13:151-156.

Cooper, A., and D. Penny. 1997. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. Science 275:1109-1113.

Cracraft, J. 2001. Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. Proceedings of the Royal Society of London B Biological Sciences 268:459-469.

Cracraft, J., F. K. Barker, M. Braun, J. Harshman, G. J. Dyke, J. Feinstein, S. Stanley, A. Cibois, P. Schikler, P. Beresford, J. García-Moreno, M. D. Sorenson, T. Yuri, and D. P. Mindell. 2004. Phylogenetic relationships among modern birds (Neornithes):

Toward an avian tree of life. Pages 468-489 *in* Assembling the Tree of Life (J. Cracraft, and M. J. Donoghue, eds.). Oxford University Press, New York.

Cubo, J. 2003. Evidence for speciational change in the evolution of ratites (Aves : Palaeognathae). Biological Journal of the Linnean Society 80:99-106.

Cutler, D. J. 2000a. Dating5. Available from the author.

Cutler, D. J. 2000b. Estimating divergence times in the presence of an overdispersed molecular clock. Molecular Biology and Evolution 17:1647–1660.

de Juana, E. 1997. Family Pteroclidae (Sandgrouse). Pages 30–49 *in* Handbook of the Birds of the World, Vol. 4, Sandgrouse to Cuckoos (J. del Hoyo, A. Elliott, and J. Sargatal, eds.). Lynx Edicions, Barcelona.

Donoghue, P. C. J., and M. J. Benton. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. Trends in Ecology & Evolution 22:424-431.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4:e88.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Dyke, G. J. 2001. The evolutionary radiation of modern birds: systematics and patterns of diversification. Geology Journal 36:305-315.

Dyke, G. J., and G. Mayr. 1999. Did parrots exist in the Cretaceous period? Nature 399:317-318.

Dyke, G. J., and M. van Tuinen. 2004. The evolutionary radiation of modern birds (Neornithes): reconciling molecules, morphology and the fossil record. Zoological Journal of the Linnean Society 141:153–177.

Easteal, S. 1999. Molecular evidence for the early divergence of placental mammals. BioEssays 21:1052 - 1058.

Edwards, S. V., and P. Beerli. 2000. Perspective: gene divergence, population divergence, and the variance in coalescent time in phylogeographic studies. Evolution 54:1839–1854.

Eldredge, N., and S. J. Gould. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Pages 82-115 *in* Models in Paleobiology (T. J. M. Schopt, ed.) Freeman, Cooper, and Co., San Francisco.

Ericson, P. G. P., C. L. Anderson, T. Britton, A. Elzanowski, U. S. Johansson, M. Källersjö, J. I. Ohlson, T. J. Parsons, D. Zuccon, and G. Mayr. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biology Letters 4:543-547.

Fain, M. G., and P. Houde. 2004. Parallel radiations in the primary clades of birds. Evolution 58:2558–2573.

Feduccia, A. 1995. Explosive evolution in tertiary birds and mammals. Science 267:637-638.

Feduccia, A. 1999. The Origin and Evolution of Birds, 2nd edition. Yale University Press, New Haven.

Feduccia, A. 2003. 'Big bang' for tertiary birds? Trends in Ecology & Evolution 18:172-176.

Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package), version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Foote, M., J. P. Hunter, C. M. Janis, and J. J. Sepkoski Jr. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of Eutherian mammals. Science 283:1310-1314.

Foote, M., and J. J. Sepkoski Jr. 1999. Absolute measures of the completeness of the fossil record. Nature 398:415-417.

Fountaine, T. M., M. J. Benton, G. J. Dyke, and R. L. Nudds. 2005. The quality of the fossil record of Mesozoic birds. Proceedings of the Royal Society of London B Biological Sciences 272:289-294.

García-Moreno, J. 2004. Is there a universal molecular clock for birds? Journal of Avian Biology 35:465-468.

Gibb, G. C., O. Kardailsky, R. T. Kimball, E. L. Braun, and D. Penny. 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Molecular Biology and Evolution 24:269-280.

Gill, F., and M. Wright. 2006. Birds of the World: Recommended English Names. Princeton University Press, Princeton, New Jersey.

Griffiths, C. S. 1994. Monophyly of the Falconiformes based on syringeal morphology. Auk 111:787-805.

Hedges, S. B., P. H. Parker, C. G. Sibley, and S. Kumar. 1996. Continental breakup and the ordinal diversification of birds and mammals. Nature 381:226-229.

Ho, S. Y. W. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. Journal of Avian Biology 38:409-414.

Ho, S. Y. W., M. J. Phillips, A. J. Drummond, and A. Cooper. 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early Metazoan radiation. Molecular Biology and Evolution 22:1355-1363.

Hope, S. 2002. The Mesozoic record of Neornithes (modern birds). Pages 339-388 *in* Mesozoic Birds: Above the heads of dinosaurs (L. M. Chiappe, and L. M. Witmer, eds.). University of California Press, Berkeley, California.

Hug, L. A., and A. J. Roger. 2007. The impact of fossils and taxon sampling on ancient molecular dating analyses. Molecular Biology and Evolution 24:1889–1897.

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

Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. Nature 392:917-920.

Linder, H. P., C. R. Hardy, and F. Rutschmann. 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. Molecular Phylogenetics and Evolution 35:569-582.

Livezey, B. C., and R. L. Zusi. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. Zoological Journal of the Linnean Society 149:1–95.

Lovette, I. J. 2004. Mitochondrial dating a mixed support for the "2% rule" in birds. Auk 121:1–6.

Magallón, S. A. 2004. Dating lineages: molecular and paleontological approaches to the temporal framework of clades. International Journal of Plant Sciences 165:S7-S21.

Marshall, C. R. 1999. Fossil gap analysis supports early Tertiary origin of trophically diverse avian orders: Comment. Geology 27:95.

Mayr, E. 1954. Change of genetic environment and evolution. Pages 157-180 *in* Evolution as a Process (J. Huxley, A. C. Hardy, and E. B. Ford, eds.). George Allen and Unwin, London.

Mayr, G. 2002. Osteological evidence for paraphyly of the avian order Caprimulgiformes (nightjars and allies). Journal für Ornithologie 143:82-97.

Mayr, G. 2005. The Paleogene fossil record of birds in Europe Biological Reviews 80:515-542.

Mayr, G., and J. Clarke. 2003. The deep divergences of neornithine birds: a phylogenetic analysis of morphological characters. Cladistics 19:527-553.

Mindell, D. P., A. Knight, C. Baer, and C. J. Huddleston. 1996. Slow rates of molecular evolution in birds and the metabolic rate and body temperature hypotheses. Molecular Biology and Evolution 13:422-426.

Mindell, D. P., J. W. Sites Jr., and D. Graur. 1989. Speciational evolution: a phylogenetic test with allozymes in Sceloporus (REPTILIA). Cladistics 5:49-61.

Mindell, D. P., M. D. Sorenson, C. J. Huddleston, H. C. J. Miranda, A. Knight, S. J. Sawchuk, and T. Yuri. 1997. Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. Pages 213–247 *in* Avian Molecular Evolution and Systematics (D. P. Mindell, ed.) Academic Press, San Diego.

Müller, J., and R. R. Reisz. 2005. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. BioEssays 27:1069-1075.

Padian, K., and L. M. Chiappe. 1998. The origin and early evolution of birds. Biological Reviews 73:1-42.

Pagel, M., C. Venditti, and A. Meade. 2006a. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science 314:119-121.

Pagel, M., C. Venditti, and A. Meade. 2006b. Test for Punctuational Evolution and the Node-Density Artifact. Available at http://www.evolution.reading.ac.uk/pe/index.html.

Paton, T., O. Haddrath, and A. J. Baker. 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proceedings of the Royal Society of London B Biological Sciences 269:839-846.

Paton, T. A., A. J. Baker, J. G. Groth, and G. F. Barrowclough. 2003. RAG-1 sequences resolve phylogenetic relationships within Charadriiform birds. Molecular Phylogenetics and Evolution 29:268-278.

Penny, D., and M. J. Phillips. 2004. The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution? Trends in Ecology & Evolution 19:516-522.

Pereira, S. L., and A. J. Baker. 2006. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Molecular Biology and Evolution 23:1731-1740.

Rambaut, A., and A. J. Drummond. 2006. TreeAnnotator, version 1.4.5. Available at http://tree.bio.ed.ac.uk/software/treeannotator/.

Rannala, B., and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645-1656.

Rannala, B., and Z. Yang. 2007. Inferring speciation times under an episodic molecular clock. Systematic Biology 56:453-466.

Rest, J. S., J. C. Ast, C. C. Austin, P. J. Waddell, E. A. Tibbetts, J. M. Hay, and D. P. Mindell. 2003. Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. Molecular Phylogenetics and Evolution 29:289-297.

Robertson, D. S., M. C. McKenna, O. B. Toon, S. Hope, and J. A. Lillegraven. 2004. Survival in the first hours of the Cenozoic. Geological Society of America Bulletin 116:760-768.

Rutschmann, F. 2006. Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. Diversity and Distributions 12:35-48.

Sanders, K. L., and M. S. Y. Lee. 2007. Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. Biology Letters 3:275-279.

Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Molecular Biology and Evolution 14:1218-1231.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 19:101-109.

Sanderson, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301-302.

Sibley, C. G., and J. E. Ahlquist. 1990. Phylogeny and Classification of Birds. Yale University Press, London.

Slack, K. E., F. Delsuc, P. A. McLenachan, U. Arnason, and D. Penny. 2007. Resolving the root of the avian mitogenomic tree by breaking up long branches. Molecular Phylogenetics and Evolution 42:1-13.

Smith, A. B., and K. J. Peterson. 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.

Sorenson, M. D., J. C. Ast, D. E. Dimcheff, T. Yuri, and D. P. Mindell. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution 12:105-114.

Sorenson, M. D., E. Oneal, J. García-Moreno, and D. P. Mindell. 2003. More taxa, more characters: the hoatzin problem is still unresolved. Molecular Biology and Evolution 20:1484-1498.

Stamatakis, A., T. Ludwig, and H. Meier. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21:456-463.

Stidham, T. A. 1998. A lower jaw from a Cretaceous parrot. Nature 396:29-30.

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

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4 (beta 10). Sinauer Associates, Sunderland, Massachusetts.

Tavaré, S., C. R. Marshall, O. Will, C. Soligo, and R. D. Martin. 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. Nature 416:726-729.

Thorne, J. L. 2003. MULTIDISTRIBUTE. Available from the author (http://statgen.ncsu.edu/thorne/multidivtime.html).

Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation. Systematic Biology 51:689-702.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of evolution. Molecular Biology and Evolution 15:1647-1657.

van Tuinen, M., and S. B. Hedges. 2001. Calibration of avian molecular clocks. Molecular Biology and Evolution 18:206-213.

van Tuinen, M., T. A. Stidham, and E. A. Hadly. 2006. Tempo and mode of modern bird evolution observed with large-scale taxonomic sampling. Historical Biology 18:205-221.

Welch, J. J., and L. Bromham. 2005. Molecular dating when rates vary. Trends in Ecology & Evolution 20:320-327.

Wills, M. A. 2007. Fossil ghost ranges are most common in some of the oldest and some of the youngest strata. Proceedings of the National Academy of Sciences of the United States of America 274:2421–2427.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in BioSciences 13:555-556.

Yang, Z., and B. Rannala. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Molecular Biology and Evolution 23:212-226.

Part II. Statistical Arbitration of Competing Molecular Phylogenetic Models

Chapter 4

Treatment of branch length parameters in partitioned phylogenetic models: an empirical case study with New World Vultures (Aves: Cathartidae)

ABSTRACT

Partitioned substitution models for phylogenetic inference are now ubiquitous in studies utilizing multiple character classes (e.g. genomes, genes, coding/non-coding regions, codon positions, stems/loops, etc.). The appeal of this approach is that heterogeneity in the evolutionary process across sundry partitions can be more finely and accurately modelled than under a single compromised model of molecular substitution. Published studies using various model selection criteria statistically justify this practice by indicating that the additional substitution model parameters required in partitioned-models leads to a significantly better fit to empirical data. However, the accommodation of branch length (BL) heterogeneity across partitions in such models has received little attention. The default method for many phylogenetic reconstruction programs supporting partition-specific rates is to treat BLs as proportional across partitions. This is a very economical way to accommodate heterogeneity across P partitions, as only P extra BL parameters are required (relative rate parameters), as opposed to an extra $P\times(2T - 3)$ parameters when estimating BLs uniquely for each partition (where T = number of taxa). Although the difference in the number of estimable parameters can vary substantially across BL treatments (depending on both P and T), the assumption of BL proportionality is rarely tested, and it is unclear how violation of this assumption might influence phylogenetic inference. We present here a molecular phylogenetic analysis of all extant species of the New World Vultures (Aves: Cathartidae) and show that while mitochondrial (mt)DNA fails to reject the proportionality assumption, support for the assumption from nuclear (nuc)DNA depends on the model selection criteria employed.

However, proportionality of BLs across the two genomes is rejected by most criteria. The optimal model for the combined alignment selected using Bayes factors is one in which BLs are constrained to be proportional for mtDNA partitions but unconstrained for nucDNA partitions. All models recover a basal split in Cathartidae between the large condor-like birds and the remaining taxa. However, relationships within these basal clades were sensitive to the model assumed. Compared to concatenated models, optimally-selected partitioned models indicate that yellow-headed vultures, once considered conspecific, are not each others closest relative; instead, the Greater Yellow-headed Vulture (*Cathartes melambrotus*) is inferred to be sister to the Turkey Vulture (*C. aura*), with the Lesser Yellow-headed Vulture (*C. burrovianus*) being sister to this novel clade. Phylogenetic placement of the California Condor (*Gymnogyps californianus*) is less clear; while linked-BL models place it sister to the King Vulture (*Sarcoramphus papa*) with high posterior probability support (0.92-1.0), relaxing the BL-proportionality assumption decreases support for this relationship (0.56-1.0), although it remains the most probable. Alternative BL-priors, while greatly influencing overall tree length and model fit, had no noticeable influence on topological reconstruction or node support values. Model fit (as gauged through the harmonic model likelihood) was much more sensitive to BL-prior for mtDNA than nucDNA, suggesting that partition-specific BL-priors may be a useful extension to constructing partitioned phylogenetic models. Inference of a time-calibrated phylogeny using partition-specific relaxed molecular clocks reveals why cathartid relationships are difficult to reconstruct: the crown clade is inferred to be ~14 million years (MY) old, and branch times for the taxa involved in the novel relationships above occurred within very short periods of time. The inferred time for the origin of the lineage is ~69 MY, consistent with the oldest putative fossil representative of the family.

INTRODUCTION

Model-based inference is ubiquitous to myriad fields of science, but perhaps has not elsewhere effected such a boon as when applied to the inherently noisy data involved in studies of ecology and evolutionary biology. An argument justifying this general approach to biological inference borders on the superfluous: i) because of the intrinsic

flexibility of modelling, finely-tuned hypotheses can be rigorously tested through a thoughtful consideration of parameterization; ii) noise/uncertainty can be explicitly accounted for and quantified; iii) model choice can be justified statistically through various model arbitration approaches; and, iv) although interpreted in different ways, these models can be applied in both maximum likelihood (ML) and Bayesian frameworks. Today, nearly all phylogenetic reconstructions are based upon explicit models of how character states change along an evolutionary trajectory, and this has heralded the unprecedented extraction of evolutionary information molecular genetic sequences.

However, devising the appropriate level of substitution model sophistication/realism is a conflicting task. On the one hand, it is clear that one should not endeavour to 'fit an elephant' (Steel, 2005), that is, try to parameterize 'reality' (the historical process of molecular substitution) in its entirety. There are a number of reasons for this: i) although accuracy may increase with the number of model parameters, so too will the variance of resulting inferences, potentially thwarting the arbitration of competing hypotheses; ii) overparameterization risks nonidentifiability (Rannala, 2002), where different combinations of parameter values generate the same likelihood, making it impossible to determine their true values; and iii) thoughtless model inflation is a form of 'data-dredging' (Burnham and Anderson, 2002), with the possible result that spurious stochastic signals (due, for example, to sampling artefacts) might be interpreted as genuine historical information. From this perspective, it thus seems reasonable to economize models as much as possible by focussing on a small number of salient components (parameters) of a given process.

On the other hand, however, there is no doubt that the standard substitution models currently in use are overly simplistic, evinced by the fact that, using a variety of model selection criteria, most typical present-day data matrices (in terms of breadth of taxon sampling and tree depth) statistically substantiate use of the general time reversible (GTR) model (Lanave et al., 1984), the most highly-parameterized substitution model in regular use. Indeed, the situation is so pervasive that the popular phylogenetics program RAxML (Stamatakis, 2006) does not even implement simpler alternative models. If

typical empirical data matrices are 'maxing-out' at GTR, this suggests that there is an overabundance of historical signal available, and that more sophisticated (realistic) models could be supported by the same data. Moreover, as the costs of molecular genetic sequencing continue to plummet, enormous amounts of data will soon be available for interrogation of increasingly subtle, but potentially pertinent, signatures of the evolutionary process. An increase in model realism (parameterization) thus appears warranted, although such extensions should of course be tempered through rigorous statistical scrutiny.

Although numerous extensions to the standard GTR are possible (e.g. time irreversible, nonhomogeneous, and nonstationary models), the most promising and direct avenue of development focuses on the heterogeneity in the substitution process across sites within an alignment. Most molecular substitution models assume that all sites are independent and identically-distributed (i.i.d.), that is, that all sites evolve according to the same evolutionary process and that observed character states are independent realizations of that process. However, it is readily apparent from even a cursory examination of different character classes (e.g. genomes, genes, coding/non-coding regions, codon positions, stems/loops, etc.) that this is not the case – for example, genes can differ greatly in GC content depending on function or genome of origin, and codon positions within a single gene often differ in the level of polymorphism by an order of magnitude. Fortunately, tools are available to accommodate such variability. For example, Yang's (1993; 1994) introduction of the one parameter discretized gamma-distributed rate heterogeneity across sites has proven indispensible to the accuracy of phylogenetic inference, and is a staple constituent of any phylogenetic analysis.

Partitioned-models (e.g. Nylander et al., 2004) offer a more general solution to relaxing the i.i.d. assumption, as heterogeneity can be accommodated not only in overall rates of substitution, but also in the substitution model itself, including individual substitution model parameter values (as well as their inclusion/exclusion) and equilibrium character state frequencies. The benefits to such a tailored approach to phylogenetic reconstruction is unassailable: more evolutionary information can be extracted, and competing partitioned-models can be rigorously arbitrated using existing statistical tools

to guard against overparameterization. The alternative and more traditional single model approach (sometimes referred to as a 'concatenated model') is thus best considered a 'compromised' model, where evolutionary signals are diluted through averaging across sundry character classes. For example, all else being equal, for two genes of equal length with respective GC-contents of 25% and 75%, a single model approach would infer a compromised GC-content of 50%; clearly an inferior parameterization for both genes. Similar compromises will involve every parameter in the model (character state frequencies, substitution model parameters, and rate heterogeneity parameters). [Strictly speaking *all* models are compromised because parameterization complexity will always fall short of reality; we thus more narrowly define 'compromised' models as those whose underparameterization relating to heterogeneity across sites can be statistically rejected by the data in hand in favour of more parameter-rich alternatives]. Empirical studies ubiquitously indicate that the relatively small number of additional substitution model parameters required in partitioned-models is statistically justified by a number of criteria such as AIC, Bayes factors, decision theory, BIC, etc. (e.g. McGuire et al., 2007). These results, together with the availability of numerous fast and flexible software packages, has resulted in the routine use of partitioned substitution models. However, despite the indisputable advantage to accommodating heterogeneity in the substitution process, the treatment of branch length (BL) parameters in such models has not received much attention (but see Marshall, 2010; Marshall et al., 2006; Pupko et al., 2002). In this paper we turn our attention to the suitability of conventional compromised BL models when analyzing partitioned molecular genetic data.

Although (as generally practiced) the number of substitution model parameters increases roughly linearly with the addition of partitions, there are three general ways to handle the number of BL parameters (n_{BL}) when analyzing data from P partitions for T taxa (Figure 4.1). First, at the one extreme, BLs can be assumed to be equal for each partition. Here, n_{BL} is simply the number of branches in an unrooted tree, $2T$-3. While this modelling has the advantage that no further BL parameters are required with added partitions (i.e. n_{BL} is independent of P), the assumption is clearly inappropriate for standard molecular partitions (e.g. codon positions, containing both hyper-variable third positions and near-invariant second positions) and can not be expected to hold on a tree

of even modest taxon sampling; given it's *a priori* unsuitability, we will not consider this modelling here. Second, BLs can be assumed to be proportional across partitions, which is the default approach taken by programs that accommodate partition-specific rates (e.g. RAxML, MrBayes). Here, a single BL vector is estimated/sampled, along with partition-specific relative-rate parameters (linked not only to rate but also the length of the respective partitions, since BLs are expressed in terms of expected numbers of substitutions per site). We will refer to this modelling strategy hereafter as 'linked-BL' models. The appeal of this modelling is its statistical economy: BL heterogeneity is accommodated through the addition of just P extra parameters (i.e. a total of $n_{BL} = (2T-3)+P$). However, this modelling represents a very strong assumption (hereafter, the 'proportionality assumption') regarding molecular evolution which entails a rather dubious biological mechanistic interpretation: that the pattern of proportional BL heterogeneity across a tree is identical for all partitions (Figure 4.1B). It is not presently clear whether such a proportionality assumption is generally supported by empirical data, especially for large trees (or large P), and routinely it remains untested. Finally, at the other extreme, BLs can be assumed to be independent (or 'unlinked') across partitions (despite being tied to the same topology) requiring a separate BL vector for each partition (hereafter, 'unlinked BL' models; Figure 4.1C). This modelling requires a total of $n_{BL} = P\times(2T-3)$, which, with fixed taxon sampling, increases very quickly with P (Figure 4.2). While this is undoubtedly more close to the reality of molecular substitution, it is unclear whether i) sufficient signal is present in a given data set to reject the proportionality assumption in favour of a substantially more parameter-rich non-proportional BL model, and ii) whether the proportionality assumption is *importantly wrong* (Box, 1976), i.e. that acceptance of the false assumption biases resulting inferences.

We applied the latter two BL modelling schemes to the phylogenetic reconstruction of the New World Vultures (Aves: Cathartidae), with the goal of identifying an optimal modelling of heterogeneity across not only genes but also genomes (nuclear and mitochondrial). Cathartidae is a small family of large-bodied carrion feeders that occur throughout North and South America. There are currently seven recognized extant cathartid species organized into five genera, although the family previously enjoyed greater diversity which included several Old World representatives in the mid- to late-

Paleogene (Rich, 1983), rendering the common name of the family a misnomer. The monophyletic status of Cathartidae has never been seriously contested. Although several previous phylogenetic analyses of large genetic matrices have involved cathartid exemplars (e.g. Brown et al., 2008; Ericson et al., 2006; Hackett et al., 2008), these studies have invariably been concerned with the relationship of Cathartidae to other avian families, rather than relationships within the family itself. For example, it is a standing question in avian systematics whether Cathartidae truly belongs within its traditional order (the diurnal raptors, Falconiformes), although concerns over the monophyletic status of the order itself complicates things (see review in Brown and Mindell, 2009). Regardless, the family-level focus of these studies has meant that Cathartidae has been poorly sampled, with the result that the enigmatic cathartid relationships have been largely overlooked, despite the fact that the most recent taxonomic revision of the family by Wetmore (1964) was more than a half century ago (but see Amadon, 1977; Rich, 1983).

METHODS

Molecular sequence data

Two mitochondrial (mtDNA) protein-coding genes (CYTB and ND2) and 5 nuclear (nucDNA) introns (EEF, GADPH, HMG, RHOD, and TGBf2) were sequenced for all (*n* = 7) recognized extant species of Cathartidae as well as 27 outgroup taxa from across Neognathae (Table 4.1). Protein-coding mtDNA was aligned at the amino acid level, while nuclear introns were aligned with ClustalW (Thompson et al., 1994) with standard gap penalties as implemented in BioEdit version 7.05 (Hall, 1999). From the original matrix, 440 alignment positions could not be aligned unambiguously (usually with respect to the Galloanserae outgroups) and so were excluded from subsequent analyses, yielding a final matrix of 5721 bp (Table 4.2).

Phylogenetic inference, partitioned-models, and the treatment of BL parameters

Phylogenetic inference was conducted using both Bayesian and ML approaches. Several partitioning strategies were considered for the complete matrix. To avoid ‘data-dredging’ (Burnham and Anderson, 2004), all candidate partition-models are based upon

empirically recognized natural genetic demarcations; these represent an *a priori* modelling of the possible congruence/dissonance in molecular substitution trajectory signatures across genomes, genes, and codon positions (the later only for mtDNA genes). We therefore did not consider, for example, 50 possible ways to partition the 5 nuclear genes, as there is no *a priori* knowledge to suggest doing this. [The number of possible ways to partition *n* elements is given by the associated Bell number; for 5 genes there are $B_n = 52$ possible partitions. We considered only two of these: all genes share the same model, or all genes have distinct models]. These constituent partitions we work with here show considerable heterogeneity in the number of variable sites present, empirical base frequencies, and (to a lesser extent) the best-fit substitution model (Table 4.2; see below). Separate analyses were performed on mtDNA-only and nucDNA-only submatrices to investigate whether the two genomes were better modelled differently. For each partitioning scheme we ran duplicate analyses: i) assuming the proportionality assumption, and ii) relaxing this assumption.

Maximum likelihood phylogenetic inference was conducted using RAxML version 7.2.0 (Stamatakis, 2006). All partitions were assigned individual GTR+I+G substitution models, which was a slight overparameterization for some the nucDNA partitions (as determined through AIC model-fitting in MrModeltest (Nylander, 2002) and PAUP* version 4 beta 10 (Swofford, 2003); Table 4.2). However, the number of 'extra' substitution model parameters involved for these partitions pales in comparison to the number of BL parameters added (see below). While conventional substitution model selection has been shown to be relatively insensitive to the particular topology employed (e.g. Abdo et al., 2005), this has not yet been demonstrated for partitioned model selection. This may be a concern if different partitioning schemes support substantially different optimal topologies. We therefore performed partitioned-model arbitration using a tree topology inferred from an unpartitioned (i.e. 'concatenated') analysis in RAxML to guard against potentially biasing partitioned-model selection towards more highly parameterized models (see below).

Bayesian partitioned-model analyses were performed in MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003), with individual partitions assigned substitution

models determined through AIC analyses in MrModeltest (Nylander, 2002) and PAUP* version 4 beta 10 (Swofford, 2003; Table 3). All substitution model parameters were unlinked across partitions, and rates across partitions were allowed to vary. To ensure a thorough sampling of parameter values from the posterior distribution, each combination of partitioning strategy, BL treatment, and BL-prior was run for 3 independent replicate MCMC analyses of 5 x 10^7 generations, with burnin set very conservatively set to 1 x 10^7 generations. Convergence of the Markov chains as well as effective sample sizes (ESS) of parameter estimates were monitored from combined MCMC log files using Tracer version 1.4 (Rambaut and Drummond, 2007).

Upon near-completion of the present study, we became aware of problems concerning BL-priors in Bayesian phylogenetic inference (Brown et al., 2010; Marshall, 2010). Briefly, these authors have demonstrated that the default BL-prior in MrBayes can lead to the inference of excessively long trees (as compared to trees inferred under ML). By default, MrBayes employs exp(λ) priors for BLs, where λ is the exponential rate parameter. The mean of an exponential distribution is 1/λ, so larger values of λ place more prior weight on shorter BLs. Here we consider three alternative exponential BL-priors to investigate a possible influence of prior on ultimate model selection: exp(10) (the default prior used in MrBayes), exp(20), and exp(100).

Model selection based on the maximized joint likelihood

To assess the relative fit of alternative candidate partitioned-models in a likelihood framework, we calculated AIC scores as:

$$\mathrm{AIC_i} = -2\log(L_i) + 2K_i,$$

where L_i is the maximized joint likelihood of model i with K_i estimable (free) parameters (Akaike, 1973). AIC is an estimate of the relative K-L discrepancy (i.e. the information lost) when using a simple model to approximate a more general model (Kullback and Leibler, 1951). Thus, the candidate model which minimizes AIC is taken as the best approximating model. Burnham and Anderson (2004) advocate the use of an alternative form of AIC which incorporates a small-sample bias-correction:

$$\mathrm{AIC_c} = -2\log(L_i) + 2K\left(\frac{n}{n-K-1}\right),$$

where n represents the sample size (taken here as the length of the alignment, 5721 bp). To assess the relative fit of our M candidate models one can calculate normalized Akaike weights:

$$w_{AIC\,i} = \frac{e^{-\frac{1}{2}\Delta AIC_i}}{\sum_{m=1}^{M} e^{-\frac{1}{2}\Delta AIC_m}},$$

where Δ_i is the $\mathrm{AIC_c}$ difference between model i and the best $\mathrm{AIC_c}$ model. Akaike weights give an indication of the degree of evidence in favour of model i being the best approximating model in the candidate set; these weights range from zero (no evidence) to one (complete evidence), although the values of individual weights depend entirely on the constituents of the candidate set of models (see below). Using these weights, it is possible to construct a 95% confidence set of models. Internal node support for alternative partitioning regimes was determined through 1000 replicates of a conserved-partition nonparametric bootstrapping procedure using the inferred optimal partitioning scheme in RAxML.

The maximized joint log-likelihood can also be used to guide Bayesian model selection. Schwarz (1978) developed the Bayesian information criterion (BIC):

$$BIC_i = -2\log(L_i) + K_i \log n$$

where the variables have the same meaning as above. Despite being similar in form to AIC, BIC is not based upon K-L information theory. Rather, it is an asymptotic approximation of the log marginal likelihood of the model assuming a particular form of prior on model parameters (Posada and Buckley, 2004). BIC penalizes the addition of parameters more heavily than AIC, and this difference becomes proportionately stronger with larger n. As a result, BIC tends to select simpler models than AIC (Posada and Buckley, 2004). As with AIC scores above, BIC scores can be normalized to give:

$$w_{BIC\,i} = \frac{e^{-\frac{1}{2}\Delta BIC_i}}{\sum_{m=1}^{M} e^{-\frac{1}{2}\Delta BIC_m}}.$$

However, because this calculation is based upon approximated marginal model likelihoods, the resultant value is not interpreted as a weight but rather an estimate of posterior model probability given equal model prior probabilities for all models in the candidate set. Analogous to the 95% confidence set of models with AIC, one can use BIC scores to construct a 95% credible set of models.

An alternative to the penalized-likelihood criteria above is the performance-based method of Minin et al. (2003) and Abdo et al. (2005). This approach, grounded in decision-theory, focuses on the relative abilities of alternative candidate models to estimate BL parameters (which are common to all models). If we assume that one model in the candidate set is the true generating model, then through using model *i* as an approximating model we will incur some loss (error) in the estimation of the 2*T*-3 BL parameters, calculated as the squared Euclidean distance between BL vectors, $\|B_i - B\|$. Extending this to multiple candidate models, the risk R_i for model *i* is the sum over all *j* models of the loss function weighted by the posterior probability of model *j*:

$$R_i = \sum_{j=1}^{M} \|B_i - B_j\| \frac{e^{-\frac{1}{2}BIC_j}}{\sum_{j=1}^{M} e^{-\frac{1}{2}BIC_j}},$$

where the probability of model *j* is approximated using BIC (above). The DT-best model is thus that which minimizes R_i (the minimum expected loss). As with BIC, DT tends to select simpler models than AIC. In the current study it was necessary to use weighted-average BLs for models the have BLs 'unlinked' across partitions, the weights being equal to the lengths of the respective partitions since BLs are expressed in units of expected substitution per site.

Model selection based on the marginal model likelihood

To contrast the fit of alternative partitioned-models in a fully-Bayesian framework, we calculated Bayes factors (BF) for each pairwise comparison of models *1* and *2*:

$$BF_{1,2} = \frac{f(data \mid M_1)}{f(data \mid M_2)},$$

where f (data | M_i) represents the marginal likelihood for partitioned-model i. Following (Nylander et al., 2004), we approximate marginal model likelihoods using the harmonic mean likelihood from post-burnin MCMC samples. Unlike the three alternative model selection criteria above, BFs accommodate uncertainty in all parameter estimates, including topology. BFs are interpreted following Kass and Raftery (1995): $2\ln(BF_{1,2}) < 6$ indicates positive evidence against M_2; $6 < 2\ln(BF_{1,2}) < 10$ indicates strong evidence against M_2; and $2\ln(BF_{1,2}) > 10$ is decisive evidence against M_2 in favour of M_1.

Divergence time estimation

To add a temporal perspective to the study of cathartid relationships, a time-calibrated phylogeny (hereafter 'chronogram') was estimated using the program BEAST version 1.4.8 (Drummond and Rambaut, 2007). The rate of molecular evolution was modelled assuming a lognormal distribution of uncorrelated branch-specific substitution rates (Drummond et al., 2006), with topology and divergence times being estimated simultaneously. We employed the optimal partitioned model identified above, and all data partitions were assigned unlinked AIC models (Table 4.2). [Investigation of the influence of 'linked' vs. 'unlinked' BL parameters in divergence time estimation is not possible, as these concepts are not applicable. BLs across partitions are neither independent nor 'linked' (i.e. proportional); rather they 'communicate' through the shared chronogram structure, with ultrametricity being enforced]. We allowed different loci to have a unique pattern of branch rate heterogeneity by jointly estimating separate lognormal distributions for each locus. However, because of the non-independence of mtDNA genes, all mtDNA partitions were assigned to a single lognormal distribution describing mtDNA branch rate heterogeneity, although sub-partitions (codon positions) were allowed to have unique relative rates. In order to avoid confounding intra- and super-specific rates of molecular evolution (Ho et al., 2005) we pruned our taxon matrix to the species level.

To allow for the placement of temporal constraints and increase the efficiency of chronogram reconstruction, 3 nodes were constrained to be monophyletic. These constrained nodes are justified by the results below and published results from a much larger data set (Hackett et al., 2008). First, the clade Galloanserae (containing all

Galliformes and Anseriformes) was constrained to allow placement of the temporal constraint derived from the 66 MY fossil duck *Vegavis iaai* (Clarke et al., 2005). Second, the clade uniting Procellariiformes and Sphenisciformes was constrained to allow placement of the temporal constraint derived from the 62 MY fossil penguin *Waimanu manneringi* (Slack et al., 2006). Third, the clade uniting Pandionidae and Accipitridae was constrained to allow placement of the temporal constraint derived from the 37 MY fossil osprey *Palaeocircus cuvieri* (Harrison and Walker, 1976). Because of the limited amount of material from *P. cuvieri*, we conducted chronogram inferences with and without this constraint. Internal temporal constraints were constructed using lognormal distributions. Because insufficient data exist regarding the extent of 'ghost lineages' in most lineages of birds, we employed standard distributions to all constraints with mean = standard deviation = 0.3, with the offset positioned to the age of the respective fossil. These distributions lend more credence to the fossil record than do conventional hard minimum constraints, but are broad enough to accommodate divergence time estimates from previous molecular studies.

Although several stemgroup fossil representatives exist for Cathartidae (Rich, 1983), many of these are of limited material and/or controversial taxonomic affinities; we therefore refrained from including these as calibration constraints. The age of Neognathae (the root of our tree) is currently in debate (e.g. Brown et al., 2007; Brown et al., 2008; Ericson et al., 2006). We therefore employed a broad uniform prior on the root spanning 130-70 MY; this prior allows for results concordant with either molecular- or fossil-inferred timescales. Nodes not directly involved in temporal constraints were modelled using a birth-death prior. For both temporal calibration regimes (i.e. those with and without *P. cuvieri*), four replicate analyses were run for 10^8 generations, sampling every 2.5×10^3 generations. Final analyses consisted of one run of 5×10^8 generations to confirm convergence/mixing. All post-burnin samples per temporal calibration regime were combined prior to parameter summary.

RESULTS AND DISCUSSION

Identification of an optimal partitioning strategy

Both ML and Bayesian approaches to partitioned-model arbitration identify significant phylogenetic structure within our alignment, and in general these approaches agreed upon partitioning strategies (Table 4.3, 4.4). It was not possible to construct confidence/credible sets of models, as for each combination of alignment/criterion a single model received essentially all of the weight (i.e. $w \sim 1.0$ and $2\ln(BF_{1,2}) > 10$). Because all BFs were extremely large when comparing models to the optimal model, only harmonic mean model likelihoods are presented rather than all pairwise calculated BFs. When considering only linked-BL models (which is what is typically done) for the full matrix, all 5 model selection criteria (AIC, AIC_c, BIC, DT-Risk, and BF) identify the 8-partition model (consisting of mtDNA codons and nucDNA genes) to fit the data significantly better than any alternative model. When the candidate set is extended to include both linked- and unlinked-BL models, AIC, AIC_c, and BF shift support for the 8-partition unlinked-BL model, again with unanimous support (AIC ~1.0, BF > 10), despite the great increase in the number of estimated parameters. BIC and DT-Risk, with their correspondingly stronger penalty for overparameterization, stuck with support for the linked-BL 8-partition model. The agreement between BIC and DT-Risk in this instances and those below is not surprising. Given that one model receives $w_{BIC} \sim 1$, BL estimation performance is only given non-zero weight relevant to this one optimal model; since that the optimal model estimates its own BLs with zero error, it is guaranteed to be the best DT-Risk model. However, it is interesting that BIC and BF do not agree on the optimally-selected model, as BIC was conceived as an approximation the marginal model likelihood (Schwarz, 1978). It thus appears that the priors assumed by BIC are dissimilar to those actually implemented in Bayesian phylogenetic inference.

Model selection considering mtDNA alone was unambiguous: all criteria selected the 3-partition model consisting of amalgamated 1st, 2nd, and 3rd codon positions across the two genes. The model receives almost complete support; the analogous unlinked 3-parition model is a distant second best. Thus, it appears that the proportionality assumption cannot be rejected for the mtDNA data in hand. Modelling of nucDNA, on

the other hand, was less clear. BF and AIC select the unlinked-BL 5-partition model, however AIC_c selects the simpler linked-BL 5-partition model while BIC and DT-Risk support the concatenated model. Thus, the appropriateness of the proportionality for nucDNA is unclear.

Nevertheless, consideration of the genome-specific results above suggest models not considered *a priori*: 1) an 8-partition model (partitioned as above) with mtDNA BLs linked to one another (to the exclusion of nucDNA BLs), and nucDNA BLs linked to one another (i.e. 2 BL vectors for $n_{BL} = 172$); and 2) an 8-partition model with mtDNA BLs linked to one another, and nucDNA BLs unlinked (6 BL vectors for $n_{BL} = 489$). Given the constituent genome-specific results, these 'mixed-BL' models are expected to provide a good fit to the data. Unfortunately, RAxML does not yet allow one to arbitrarily link BLs (i.e. they are either all linked or all unlinked), and thus we can not implement these models for phylogenetic reconstruction in a ML framework. However, we can approximate the fit of these models on a fixed tree topology: because in both of the mixed-BL models above all parameters (substitution model and BL parameters) are completely unlinked across genomes, when considering the same topology for all partitions one can simply sum the log-likelihoods of the two constituent sub-models as an approximation of the joint log-likelihood of the full model. Doing this, we see that AIC selects the more general mixed-BL model while AIC_c goes with the simpler mixed-BL model. BIC and DT-Risk, on the other hand, remain in support of the 8-partition linked-BL model (the simpler of the two mixed-BL models is a distant second best). MrBayes allows for implementation of the full mixed-BL models. Adding these two new models to the candidate set, BF support for the most general mixed-BL model is decisive. These results thus suggest mixed support for the proportionality assumption: while mtDNA cannot reject proportionality of BLs across partitions, proportionality of BLs across the two genomes is rejected using most criteria. The non-proportionality of BLs across genomes is readily apparent from an examination of genome-specific BLs estimated on a single topology (Figure 4.4).

Several authors have recently commented on the connection between inappropriate BL-priors and the inference of excessively long trees (Brown et al., 2010; Marshall,

2010), measured as the sum of all 2T-3 BLs in the tree (each expressed in units of expected number of substitutions per site), although these authors find no influence of BL-prior on either topological inference or node support. We were interested here in whether priors of different breadth might ultimately lead to the selection of alternative partitioned models, especially since we explicitly consider the treatment of BL parameters in partitioned model inference. As expected, narrower exponential priors (i.e. those with larger exponential rate parameters, λ) lead to the inference of shorter trees, although this tended to be much more pronounced for mtDNA than nucDNA (Table 4.4). For mtDNA, tree lengths often differed by a factor of two across the various priors. We also note an influence of BL-prior on the degree of model fit as gauged from the harmonic mean model likelihood. While analyses of nucDNA alone showed very little influence of BL-prior on model likelihoods, mtDNA-only analyses experienced considerable changes, especially with unlinked-BL models.

This can be understood through the respective forms of the prior and the likelihood (Figure 4.3). If all parameter values of non-negligible likelihood are contained well within the boundaries of the prior (in the case of the exponential, close to zero), then constricting the breadth of the prior should lead to a superior marginal likelihood because regions of small likelihoods in the original prior need not be integrated over (in the case the case of the exponential, these extreme values are down-weighted, rather than excluded as would be the case in a uniform distribution). Because mtDNA marginal likelihoods decrease so much with narrower priors, this suggests that the highest likelihood BL parameter values for these sites lie within the tail of the narrower priors. Indeed, ML BL estimates for from RAxML (Figure 4.4) reveal that these values are extreme when considering the strongest BL-prior ($\lambda = 100$; mean = 0.01), in particular for mtDNA. Thus, decreases in marginal model likelihoods with changes in BL-priors for the full matrix appear to be due primarily to the poorer fit for mtDNA. These findings suggest another direction of research: investigating the relative fit of candidate models while attaching distinct BL-priors to individual partitions. Nevertheless, form of BL-prior had no influence on which partitioning scheme was ultimately selected as the optimal strategy (Table 4.4).

Phylogenetic reconstruction of Cathartidae – are our models 'importantly wrong'?

Given the vast superiority of the optimal models compared to the alternatives, the natural question to ask is: does it matter? Do models that better fit the data return unique solutions that alternative models are not able to (perhaps to under- or mis-parameterization)? Empirical phylogeneticists are ultimately interested in two attributes: the optimal tree topology returned, and node bi-partition support (either nonparametric bootstrap proportions or posterior node probabilities).

In general, all models considered here reveal a basal split within extant Cathartidae of two clades with unanimous support in both ML and Bayesian implementations: i) the genus *Cathartes* [Lesser- (*burrovianus*) and Greater (*melambrotus*) Yellow-headed Vultures, and Turkey Vulture (*aura*)] together with the Black Vulture (*Coragyps atratus*), and ii) the Andean (*Vultur gryphus*) and California (*Gymnogyps californianus*) Condors, and the King Vulture (*Sarcoramphus papa*). However, relationships within these two clades were not insensitive to model choice.

Figure 4.5 contrasts trees inferred assuming simple versus optimal models. Both ML and Bayesian approaches employing a simple single-partition model recover a monophyletic yellow-headed (YH) vulture clade, sister to the Turkey Vulture. The YH clade is quite strongly supported in the Bayesian analysis with a posterior clade probability of 0.94, although ML nonparametric bootstrap support (61%) is more modest. However, when employing the 8-partition model deemed optimal by most model selection criteria considered here, *C. melambrotus* is instead inferred to be the sister taxon of *C. aura*, with *C. burrovianus* being sister to this clade, thus rendering 'YH vultures' paraphyletic. This is particularly interesting, as from their first description in 1845 the two taxa were originally considered conspecific; only in 1964 were they raised to the species level (Wetmore, 1964). Our results suggest that the two species of YH vultures are even more distinct than Wetmore realized.

The other point of model sensitivity involves the phylogenetic placement of the California Condor. In general, most models recover a sister relationship of this taxon with the King Vulture. However, support for this relationship varies substantially with the

model employed. Interestingly, the MLE tree for the 8-partition model and unlinked BLs does not recover this relationship, although the bootstrap consensus tree for this model does recover the clade. This suggests that there may be just a few sites supporting an alternative relationship, rare enough that they are being missed in the bootstrap resampling. This interpretation is consistent with the generally young crown age of this taxon (see below).

Overall, these two relationships of interest show different patterns in terms of model-sensitivity (Table 4.5). In general, support for paraphyletic YH vultures increased when removing the constraint of proportional BLs for ML analyses, but remained essentially constant for Bayesian analyses. Support for the sister relationship between the King Vulture and California Condor instead generally decreased when assuming unlinked BLs; this effect was more pronounced for Bayesian than ML analyses. Interestingly, the two most probable trees for each of the unlinked-BL analyses differed only in the placement of the California Condor, the second most probable placement being sister to the Andean Condor. The decrease in support precision for the placement of the California Condor in more complex models may indicate that, in regards to this specific phylogenetic hypothesis, the models are overparameterized (see below). Resolution of the placement of this taxon may require further data.

Although our study was not designed to investigate the relationship of Cathartidae to other neoavian families, we can nonetheless reject a long-standing hypothesis is avian systematics. This pertains to a purported relationship between cathartid vultures with storks (Aves: Ciconiiformes). Although suggested from morphological data (Garrod, 1874; Ligon, 1967), the influence of this hypothesis on taxonomy ultimately stems from two genetic studies. Sibley and Ahlquist (1990), using now largely discredited DNA-DNA hybridization data, argued that cathartid vultures are a sub-family within Ciconiidae. The second study, using mtDNA cytochrome b sequences, suggested an alliance between some storks and some New World Vultures (Avise et al., 1994). However, it was discovered almost immediately that this second study contained erroneous sequences (Hackett et al., 1995; Helbig and Seibold, 1996), rendering the conclusions untenable. Despite having been long since retracted (Avise and Nelson,

1995), many papers continue to cite this study as if it were true (e.g. Tagliarini et al., 2009). No recent study with broad taxon sampling has been able to reconstruct a cathartid-stork relationship (e.g. Brown et al., 2008; Ericson et al., 2006; Hackett et al., 2008; Livezey and Zusi, 2007). We include in the present study two stork species, the Maguari Stork (*Ciconia maguari*) and the Saddle-billed Stork (*Ephippiorhynchus senegalensis*). In no combination of model/prior are storks and cathartid vultures even remotely related; instead, the storks cluster with the other Ciconiiformes taxa, the herons, Shoebill, and Hammerkop.

A temporal perspective on the diversification of Cathartidae

Chronogram reconstruction reveals why phylogenetic inference within Cathartidae is so difficult (Figure 4.6). Results for the ingroup taxa were insensitive to the inclusion/exclusion of the controversial 37 MY fossil osprey (*P. cuvieri*) constraint (Harrison and Walker, 1976), while deeper nodes changed < 3MY; only results with the constraint are presented. Analyses employing sophisticated relaxed molecular genetic clocks (with rate variation across lineages fitted to distinct lognormal distributions associated with the optimal partitioned model above; Table 4.6) indicates that crown Cathartidae diversified ~14 million years ago (MYA) (17.5-11.2 credible interval). The stem lineage is inferred to have originated ~69 MYA (75.8-64.3 credible interval). This is somewhat younger than previously estimated from nuclear (Brown et al., 2007) and mitochondrial (Brown et al., 2008) DNA; we attribute this to increased taxon sampling and a more nuanced relaxed molecular clock approach. The oldest putative stem-cathartid fossil material is *Paracathartes howardae* (Harrison and Walker, 1977) from Wyoming approximately 55 MYA (early Eocene). This is somewhat contentious, as the material is represented by a single tibiotarsus and affinities are controversial. However, our results indicate that this fossil at least fits within our temporal reconstruction of the family.

Diversification of the two unanticipated phylogenetic results above appears to have been rapid, with multiple lineages arising in a period of less than one million years (see overlapping credibility intervals on divergence time estimates; Figure 4.6). Given this rapid radiation, morphological congruence amongst unrelated taxa is understandable. We note that posterior clade probabilities for these phylogenetic relationships differs

somewhat from the analyses above. Given the sensitivity of these focal relationships to model choice this is not surprising, as BEAST does neither constrain BLs to be proportional nor allow them to be freely unconstrained. Rather, BLs must jointly satisfy the ultrametricity of the time-calibrated phylogeny, meaning that some BLs may be proportional across partitions while others are not. Regarding the sister relationship between King Vulture and California Condor, we see that support increases to a posterior probability of 0.94, on par with results from unlinked-BL models. As mentioned above, this may indicate that, with respect to this specific clade, unlinking BLs across partitions may constitute overparameterization. On the other hand, we see that support for the paraphyly of YH vultures decreased to 0.6, a value much lower than those generated from MrBayes (0.82-1.0; Table 4.5). This decrease in support understandable given the near-coeval inferred divergence times: *C. burrovianus* estimated to have originated 2.71 MYA (3.74-1.76 credible interval), followed by the split of *C. melambrotus* and *C. aura* just 500,000 years later (mean 2.21 MY, 3.09-1.38 credible interval).

Inferred characteristics of the partition-specific relaxed molecular clocks employed here lend credence to our optimal selection of partitioning strategy (Table 4.6). Each partition is described by a unique combination of mean rate, coefficient of variation (an indication of clock-likeness), covariance (the degree of autocorrelation in rates across ancestor-descendant branches), and standard deviation of rates estimates. MtDNA, for example, evolves at more than an order of magnitude faster than any of the nuclear genes, but is not correspondingly more unclock-like. Interestingly, nucDNA gene RHOD fails to reject an autocorrelation of rates (that is, include zero in the credible interval). In virtually every study/system that has used the UCLN model in BEAST autocorrelation is rejected: birds (Brown et al., 2008), fish (Alfaro et al., 2007), mammals (Kitazoe et al., 2007), viruses (Drummond et al., 2006), photosynthetic algae (Brown and Sorhannus, in revision), and plants (Renner et al., 2008; Zhong et al., 2009). Thus, RHOD, in failing to reject autocorrelation, appears to be idiosyncratic. This is important, because most relaxed molecular clock methods available explicitly assumed that rates are autocorrelated across a tree (Brown and van Tuinen, 2010; Rutschmann, 2006).

CONCLUSIONS

Our results indicate that assumed model complexity does indeed *matter* (Box, 1976) i.e. models of different sizes can lead to disparate inferences. We show here that alternative treatment of BL parameters in partitioned phylogenetic models can produce drastically different degrees of fit to the data. How general this is remains to be seen; however, given the potential boon to empirical phylogenetic inference, it seems prudent to consider these classes of models in our candidate sets. Finally, although influencing inferred tree length and degree of model fit (and considerably so for mtDNA), we find no influence of BL-priors on topological reconstruction of posterior clade probabilities. However, individual partitions were differently influenced by BL-prior, suggesting that future partitioned models may do well to consider partition-specific BL-priors.

ACKNOWLEDGMENTS

We would like to thank R. Kimball for nuclear intron PCR primer sequences. Funding was provided by the University of Michigan Rackham Graduate School and Museum of Zoology (JWB), the Peregrine Fund (JAJ), and the National Science Foundation (DPM). JWB thanks A. Schoenberg, G. Ligeti, and D. Van Vliet for encouragement throughout.

Table 4.1 Taxa included in the present study.

Ingroup (Cathartidae)	**Common Name**
Cathartes aura (*n* = 3)	Turkey Vulture
Cathartes burrovianus	Lesser Yellow-headed Vulture
Cathartes melambrotus	Greater Yellow-headed Vulture
Coragyps atratus (*n* = 2)	Black Vulture
Vultur gryphus (*n* = 2)	Andean Condor
Sarcoramphus papa (*n* = 4)	King Vulture
Gymnogyps californianus (*n* = 2)	California Condor

Non-cathartid Falconiformes	**Family**
Falco femoralis	Falconidae
Falco peregrinus	Falconidae
Falco rusticolus	Falconidae
Polyborus plancus	Falconidae
Gyps bengalensis	Accipitridae
Pandion haliaetus	Pandionidae
Sagittarius serpentarius	Sagittariidae

Non-falconiform Neoaves	**Family**
Asio otus	Strigidae
Apterodytes patagonicus	Spheniscidae
Pygoscelis antarcticus	Spheniscidae
Fulmarus glacialis	Procellariidae
Larus marinus	Laridae
Lunda cirrhata	Alcidae
Charadrius alexandrinus	Charadriidae
Gavia immer	Gaviidae
Ciconia maguari	Ciconiidae
Ephippiorhynchus senegalensis	Ciconiidae
Nyctanassa violacea	Ardeidae
Ardea herodias	Ardeidae
Scopus umbretta	Scopidae
Balaeniceps rex	Balaenicipitidae

Galloanserae (outgroup)	**Family**
Anas clypeata	Anatidae
Anas platyrhunchos	Anatidae
Bucephala albeola	Anatidae
Gallus domesticus	Phasianidae
Gallus gallus	Phasianidae
Bonasa umbellus	Phasianidae

Table 4.2 Fragment properties after excluding ambiguously-aligned sites.

Gene (genome)	Aligned length (bp)	Variable sites (%)	Parsimony informative sites (%)	AIC model (# parameters)	Empirical base frequencies			
					A	C	G	T
CYTB (mtDNA)	1029	498 (48.4)	439 (42.7)	GTR+I+G (10)	0.270	0.349	0.134	0.247
ND2 (mtDNA)	1037	632 (60.9)	564 (54.4)	GTR+I+G (10)	0.315	0.346	0.102	0.237
mtDNA 1st	688	308 (44.8)	261 (37.9)	GTR+I+G (10)	0.302	0.298	0.192	0.208
mtDNA 2nd	689	147 (21.3)	97 (14.1)	GTR+I+G (10)	0.182	0.314	0.110	0.394
mtDNA 3rd	689	675 (98.0)	645 (93.6)	GTR+I+G (10)	0.394	0.432	0.050	0.124
All (mtDNA)	2066	1130 (54.7)	1003 (48.5)	GTR+I+G (10)	0.293	0.348	0.118	0.242
EEF (nucDNA)	924	533 (57.7)	391 (42.3)	GTR+G (9)	0.275	0.273	0.205	0.247
GADPH (nucDNA)	416	272 (65.4)	230 (55.3)	HKY+I+G (6)	0.211	0.210	0.319	0.260
HMG (nucDNA)	673	502 (74.6)	417 (62.0)	GTR+G (9)	0.288	0.148	0.231	0.333
RHOD (nucDNA)	979	639 (65.3)	487 (49.7)	GTR+G (9)	0.208	0.240	0.283	0.269
TGFb2 (nucDNA)	663	368 (55.5)	250 (37.7)	HKY+G (5)	0.251	0.221	0.231	0.297
All (nucDNA)	3655	2331 (63.8)	1775 (48.6)	GTR+G (9)	0.248	0.224	0.248	0.280
All (mtDNA+nucDNA)	5721	3461 (60.5)	2778 (48.6)	GTR+I+G (10)	0.264	0.270	0.200	0.266

Table 4.3 Identifying the optimal partitioning strategy based on the maximized joint log-likelihood. The number of constituent partitions and the treatment of BL parameters (link/unlink) for each model is given in square brackets. For the full matrix three sets of weights (w) are presented for both AIC criteria: the topmost value uses only linked-BL models as the candidate set; the middle value considers both linked- and unlinked BL models; the lowermost value considers all candidate models. Bold values indicate optimal models per matrix/criterion.

Full Matrix Models	**-ln(L)**	**Parameters (n_{BL})**	**AIC**	w	**AIC_c**	w_c	**BIC**	w_{BIC}	**DT-Risk**
[1] All	56672.17	91 (81)	113526	0	113529	0	114132	0	0.97661
[2-link] mtDNA, nucDNA	55205.55	103 (83)	110617	0	110621	0	111302	0	0.64045
[6-link] mtDNA, nuc-genes	55089.33	147 (87)	110473	0	110480	0	111450	0	0.62650
[7-link] mt-genes, nuc-genes	55051.09	158 (88)	110418	0	110427	0	111469	0	0.60828
[8-link] mt-codons, nuc-genes	53395.23	169 (89)	107128	$\sim 1_{link}$	107139	$\sim 1_{link}$	**108253**	~1	**0**
[2-unlink] mtDNA, nucDNA	54731.28	182 (162)	109827	0	109839	0	111037	0	0.87700
[6-unlink] mtDNA, nuc-genes	54252.44	546 (486)	109597	0	109712	0	113229	0	0.87504
[7-unlink] mtDNA genes, nuc-genes	54159.79	637 (567)	109594	0	109753	0	113831	0	0.93295
[8-unlink] mtDNA codons, nuc-genes	52623.19	728 (648)	106702	$\sim 1_{link/unlink}$	106915	$\sim 1_{link/unlink}$	111544	0	0.51448
[8] link-mt-codons, link-nuc-genes	53131.06	250 (172)	106762	0	**106785**	$\sim 1_{all}$	108425	0	0.50404
[8] link-mt-codons, unlink-nuc-genes	52770.03	568 (489)	**106678**	$\sim 1_{all}$	106804	0	110463	0	0.53474
mtDNA-only Models									
[1] concat. mtDNA	25902.19	91 (81)	51986	0	51995	0	52592	0	1.13579
[2-link] mt-genes	25870.41	103 (83)	51947	0	51958	0	52632	0	1.33842
[3-link] mt-codons	24419.78	114 (84)	**49068**	~1	**49081**	~1	**49826**	~1	**0**
[2-unlink] mt-genes	25809.54	182 (162)	51983	0	52018		53194	0	1.33752
[3-unlink] mt-codons	24272.94	273 (243)	49092	0	49175	0	50908	0	1.05996
nucDNA-only Models									
[1] concat. nucDNA	28829.09	91 (81)	57840	0	57845	0	**58445**	~1	**0**
[5-link] nuc-genes	28711.28	136 (86)	57695	0	**57705**	~1	58599	0	0.01944
[5-unlink] nuc-genes	28350.25	455 (405)	**57610**	~1	57740	0	60637	0	0.23412

Table 4.4 Identifying optimal partitioned-models using marginal model likelihoods. The number of constituent partitions and the treatment of BL parameters (link/unlink) for each model is given in square brackets. BL-priors are denoted as BL-X, where X is the exponential rate parameter. T_L is the weighted-average posterior mean tree length for the particular combination of model/BL-prior. Values in bold represent the optimal partitioned-model for a given alignment/BL-prior.

		Harmonic mean model –ln(*L*)		
Full Matrix Models	**Parameters (n_{BL})**	**BL-10 (T_L)**	**BL-20 (T_L)**	**BL-100 (T_L)**
[1] All	91 (81)	56743 (2.808)	56738 (2.769)	56749 (2.527)
[2-link] mtDNA, nucDNA	103 (83)	54793 (5.626)	54785 (5.100)	54854 (3.542)
[6-link] mtDNA, nuc-genes	135 (87)	54657 (5.602)	54663 (5.083)	54793 (3.543)
[7-link] mtDNA genes, nuc-genes	146 (88)	54629 (5.713)	54633 (5.148)	54710 (3.548)
[8-link] mtDNA codons, nuc-genes	157 (89)	53444 (6.888)[1]	53448 (5.949)[1]	53538 (3.921)[1,2]
[2-unlink] mtDNA, nucDNA	182 (162)	54583 (4.337)	54591 (3.803)	54726 (2.497)
[6-unlink] mtDNA, nuc-genes	534 (486)	54360 (4.493)	54366 (3.908)	54518 (2.385)
[7-unlink] mt-genes, nuc-genes	625 (567)	54349 (4.330)	54637 (3.908)	54607 (2.054)
[8-unlink] mt-codons, nuc-genes	716 (648)	53279 (4.533)[2]	53254 (3.497)[2]	53649 (2.054)
[8] link-mt-codons, link-nuc-genes	240 (172)	53258 (4.967)	53274 (4.177)	53431 (2.816)
[8] link-mt-codons, unlink-nuc-genes	557 (489)	**53169 (5.086)**	**53176 (4.251)**	**53332 (2.729)**
mtDNA-only Models				
[1] mtDNA	91 (81)	25651 (9.017)	25661 (7.495)	25803 (4.059)
[2-link] mt-genes	103 (83)	25631 (9.093)	25642 (7.541)	25776 (4.064)
[3-link] mt-codons	114 (84)	**24457 (10.428)**	**24470 (8.395)**	**24614 (4.925)**
[2-unlink] mt-genes	182 (162)	25643 (8.651)	25667 (6.553)	25906 (3.148)
[3-unlink] mt-codons	273 (243)	24552 (8.632)	24544 (6.407)	24944 (3.147)
nucDNA-only Models				
[1] nucDNA	91 (81)	28910 (1.668)	28900 (1.656)	28901 (1.569)
[5-link] nuc-genes	124 (86)	28782 (1.665)	28785 (1.654)	28782 (1.568)
[5-unlink] nuc-genes	443 (405)	**28696 (1.866)**	**28691 (1.781)**	**28713 (1.417)**

[1] Optimal model if only linked-BL models are considered; [2] Optimal model if only linked- and unlinked-BL models are considered.

Table 4.5 ML (non-parametric bootstrap) and Bayesian (posterior clade probability) support for nodes of interest across models and branch length priors. BL-X refers to the exponential prior placed on branch length parameters, where X is the exponential rate parameter. 'A' refers to support for a paraphyletic yellow-headed vultures. 'B' refers to the node uniting the King Vulture and California Condor. It is not currently possible to implement the mixed-BL models in RAxML.

	ML Bootstrap		**BL-10**		**BL-20**		**BL-100**	
Full Matrix Models	**A**	**B**	**A**	**B**	**A**	**B**	**A**	**B**
[1] All	39	56	0.06	0.98	0.06	0.99	0.04	0.99
[2-link] mtDNA, nucDNA	47	59	0.99	0.97	0.98	097	1.0	1.0
[6-link] mtDNA, nuc-genes	47	62	0.98	0.98	0.98	0.98	0.93	0.98
[7-link] mtDNA genes, nuc-genes	63	62	0.97	0.96	0.97	0.96	0.92	0.96
[8-link] mtDNA codons, nuc-genes	68	60	0.99	0.96	0.99	0.96	0.94	0.96
[2-unlink] mtDNA, nucDNA	63	54	0.99	0.88	0.99	0.9	0.96	0.89
[6-unlink] mtDNA, nuc-genes	72	49	1.0	0.65	1.0	0.66	1.0	1.0
[7-unlink] mt-genes, nuc-genes	73	47	0.95	0.79	0.95	0.75	0.83	0.89
[8-unlink] mt-codons, nuc-genes	73	65	0.97	0.57	0.93	0.58	0.82	0.56
[8] link-mt-codons, link-nuc-genes	NA	NA	0.99	0.89	0.99	0.88	0.95	0.89
[8] link-mt-codons, unlink-nuc-genes	NA	NA	1.0	0.58	1.0	0.58	0.98	0.68

Table 4.6 Quantification of inferred rate variation amongst locus-specific relaxed molecular clocks. Results from BEAST analyses employing separate but jointly-sampled lognormal distributions to describe uncorrelated among-branch rate variation for each locus. Values represent mean and 95% highest posterior density (HPD) intervals (lower, upper). The coefficient of variation is an indication of the clock-likeness of the data (smaller values being more clock-like), while covariance is a measure of the autocorrelation of rates from ancestor to descendent branches (higher values representing more highly-autocorrelated rates). UCLN = uncorrelated lognormal.

Locus	Coefficient of variation	Covariance	UCLN mean	UCLN standard deviation
mtDNA	0.442 (0.339,0.554)	0.189 (-0.050,0.421)	1.06×10^{-2} (8.39×10^{-3},1.29×10^{-2})	0.478 (0.346,0.623)
EEF	0.362 (0.243,0.487)	0.085 (-0.154,0.320)	9.99×10^{-4} (8.56×10^{-4},1.15×10^{-3})	0.367 (0.245,0.497)
GADPH	0.573 (0.396,0.774)	0.074 (-0.170,0.328)	1.43×10^{-3} (1.13×10^{-3},1.77×10^{-3})	0.570 (0.391,0.767)
HMG	0.882 (0.585,1.248)	-0.059 (-0.232,0.132)	1.53×10^{-3} (1.06×10^{-3},2.10×10^{-3})	0.859 (0.593,1.134)
RHOD	0.827 (0.653,1.008)	0.259 (0.040,0.501)	9.91×10^{-4} (7.52×10^{-4},1.27×10^{-3})	0.821 (0.638,1.035)
TGFb2	0.449 (0.295,0.610)	0.063 (-0.173,0.302)	6.92×10^{-4} (5.66×10^{-4},8.21×10^{-4})	0.469 (0.302,0.647)

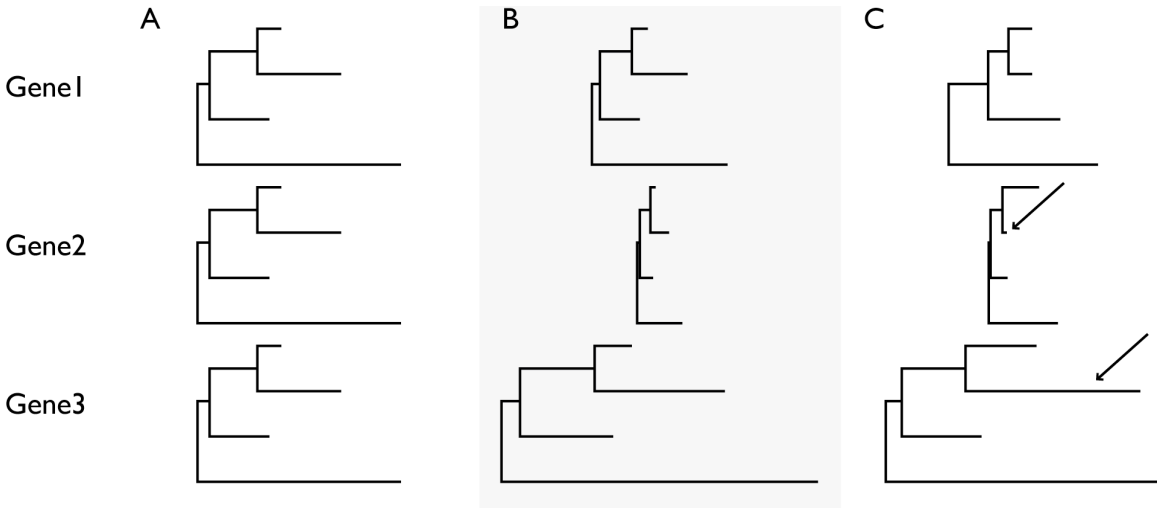

Figure 4.1 The potential treatment of branch length (BL) parameters with P partitions and T taxa. Three hypothetical genes are analyzed for 4 taxa; Gene2 is a 'slow' gene, Gene3 is a 'fast' gene, and Gene1 is intermediate. (A) Branches are assumed to be equal for every gene; $2T - 3 = 5$ total BL parameters. Although no extra BL parameters are required under this scenario, the resulting compromised BL vector is clearly inappropriate (see B-C). (B) Branches are assumed to be perfectly proportional across genes; $(2T - 3) + P = 8$ total BL parameters. Here, only 3 extra BL parameters are needed to accommodate branch length heterogeneity across genes. However, this approach requires that gene-specific phylograms differ exclusively by scale, and this assumption may not be supported by the data in hand. (C) Branch lengths (but not topology) are independent across genes. Unlike the previous two options, this model allows for heterogeneity in the pattern of BLs across genes; e.g. a relatively 'slow' lineage for Gene2 can be a 'fast' lineage for a Gene3. Despite the increased realism of this model, it requires a total of $P \times (2T - 3) = 15$ BL parameters which may not be supported with a limited data set.

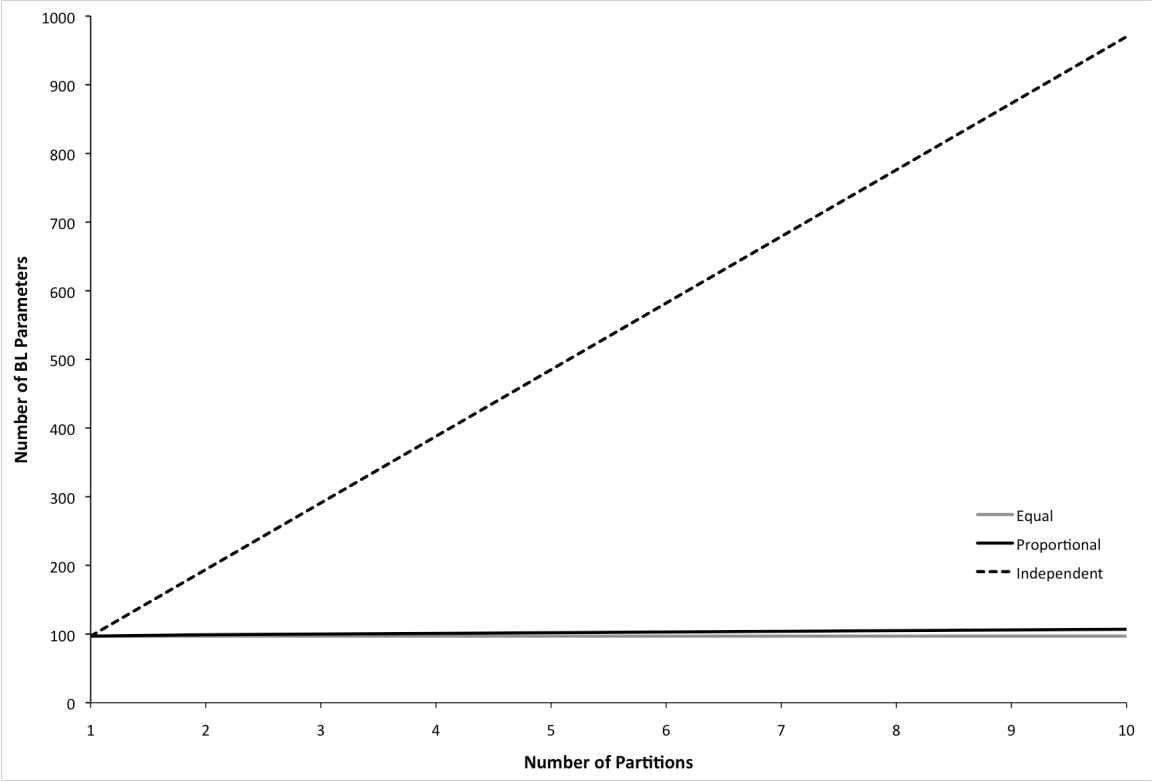

Figure 4.2 The number of branch length parameters (n_{BL}) in partitioned-models for 50 taxa and P partitions. n_{BL} is invariant to P if BLs are assumed equal for every partition. Assuming proportionality of BLs across partitions requires P extra relative-rate parameters, but only a single BL vector. An assumption of independence of BLs across partitions requires P separate BL vectors, which increases very quickly with P.

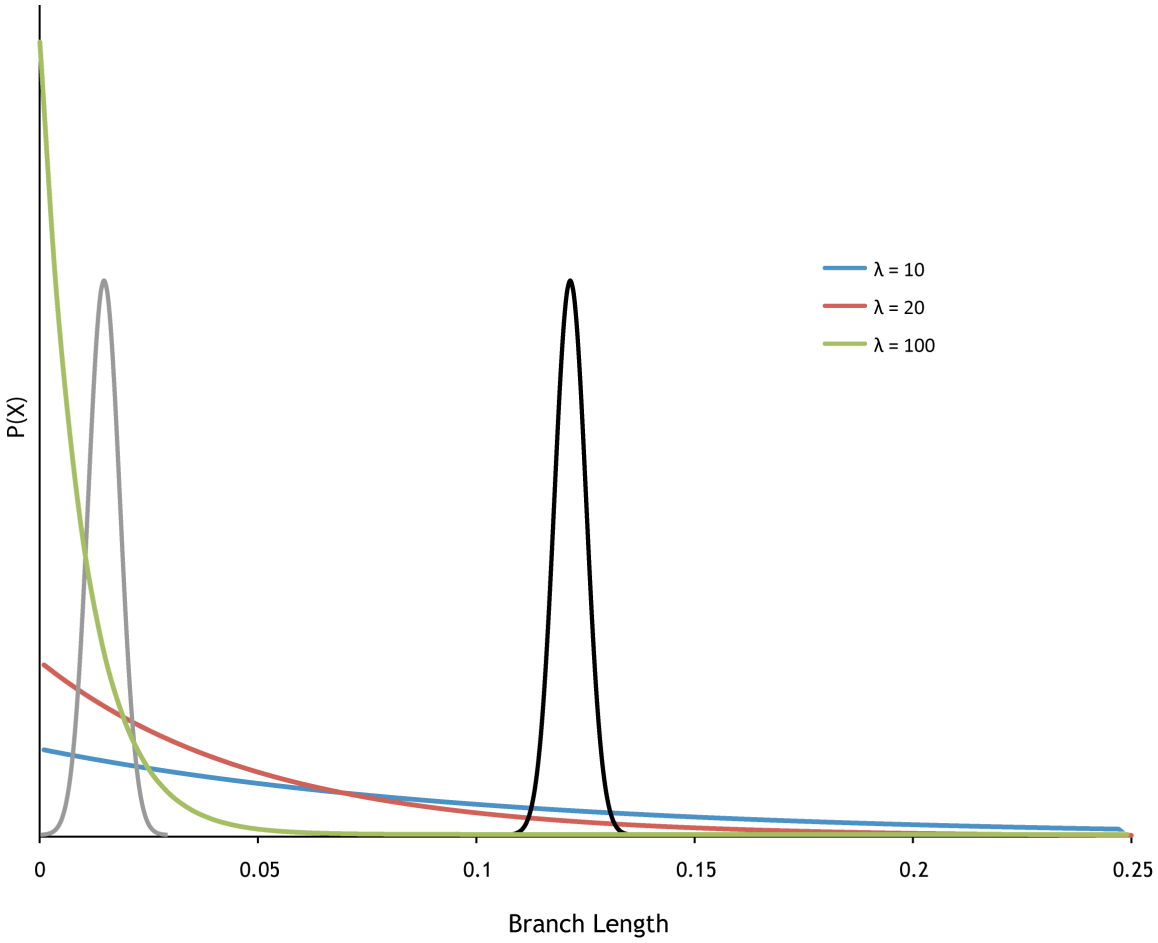

Figure 4.3 Influence of BL-priors on marginal likelihoods. The blue, red, and green curves represent exponential priors that successively place a larger volume of the prior on shorter branch lengths (i.e. larger exponential rate parameters, λ), while the black and grey curves represent hypothetical likelihood surfaces for parameter values (where values outside the curves have small but non-zero likelihoods). The marginal likelihood of parameter values is the product of the likelihood and prior, integrated over the breadth of the prior. In the case where the highest likelihood values for the parameter lie close to zero (grey curve), using a more narrow prior will generate a higher marginal likelihood, as extreme values of the parameter will contribute little to the integral since both the prior and likelihood are small. In the case where the highest likelihood values for the parameter lie instead far from zero (black curve), using a narrower exponential prior will instead tend to decrease the marginal likelihood, as small values for the parameter receive the bulk of the volume of the prior. The ultimate fate of the marginal likelihood will depend on the relative strength of the prior and the likelihood (i.e. the amount of data).

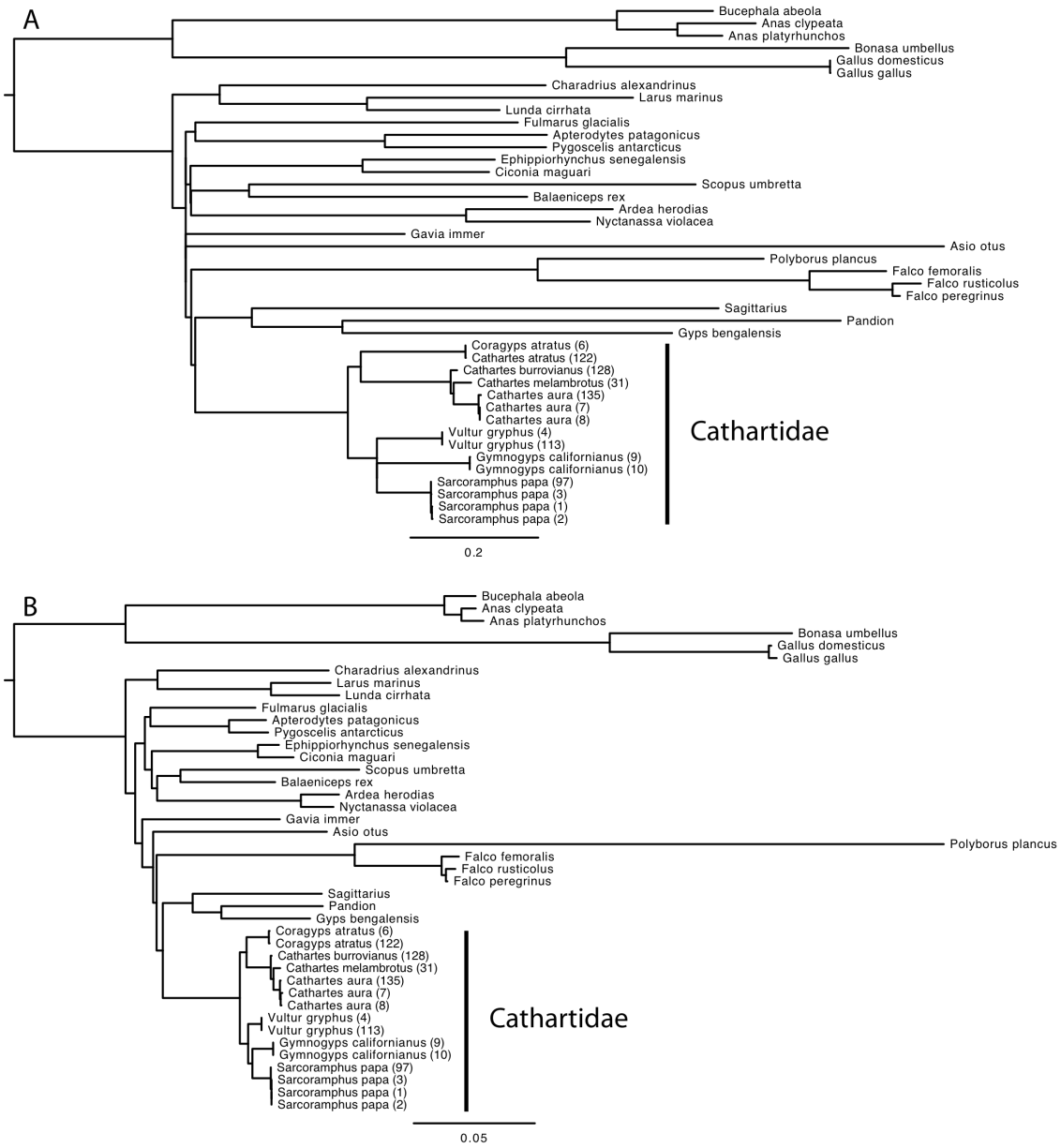

Figure 4.4 A comparison of ML branch length heterogeneity within and between genomes. (A) Phylogram estimated from mtDNA partitioning by codon position (n = 3 partitions) while enforcing BLs to be proportional across partitions. (B) Phylogram estimated from nucDNA partitioned by gene (n = 5 partitions) while not requiring BLs to be proportional across partitions; BLs shown are weighted-averages across the 5 partitions. BLs were estimated using the same tree topology for both genomes. Note differences in scale bars. mtDNA BLs tend to be much longer than analogous nucDNA BLs, and can be considered 'extreme' values when assuming an exp(100) BL-prior in Bayesian inference.

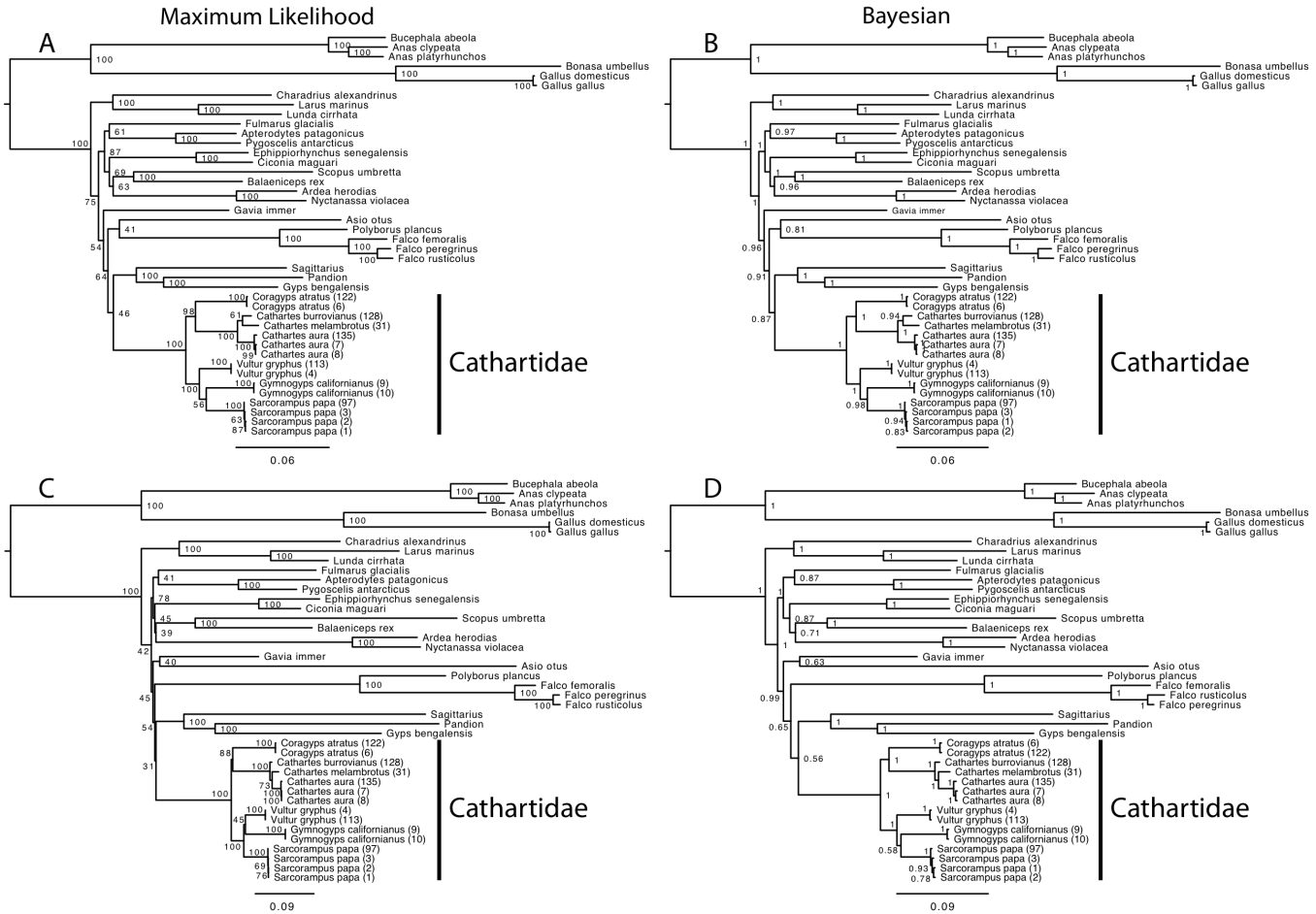

Figure 4.5 The influence of model complexity on phylogenetic inference. Phylograms (A) and (B) were inferred using a single 'concatenated' substitution model, while phylograms (C) and (D) were inferred using the inferred optimal 8-partition model for ML (A and C) and Bayesian (B and D) implementations. Partitioned ML analyses (C) implemented the model where BLs are unlinked across all partitions (the most general possibility using RAxML); node support values are from 500 nonparametric bootstrap replicates, and are mapped onto the MLE tree. Partitioned Bayesian analyses (D) implemented the 8-partition 'mixed-BL' model, where mtDNA BLs are enforced to be proportional but nucDNA BLs are unconstrained; node values are posterior clade probabilities mapped onto the majority-rule consensus tree. Both single partition models recover a monophyletic yellow-headed vulture clade (*C. melambrotus* and *C. burrovianus*) while 8-partition models reconstruct *C. burrovianus* as basal to a clade of *C. melambrotus* and the Turkey Vulture (*C. aura*). The California Condor (*Gymnogyps californianus*) tends to group with the King Vulture (*Sarcoramphus papa*), albeit with mediocre support. This support decreases with model complexity in Bayesian analyses. The MLE tree for the 8-partition model (C) does not recover this grouping, although the majority-rule consensus tree does (not shown), with support of 60%.

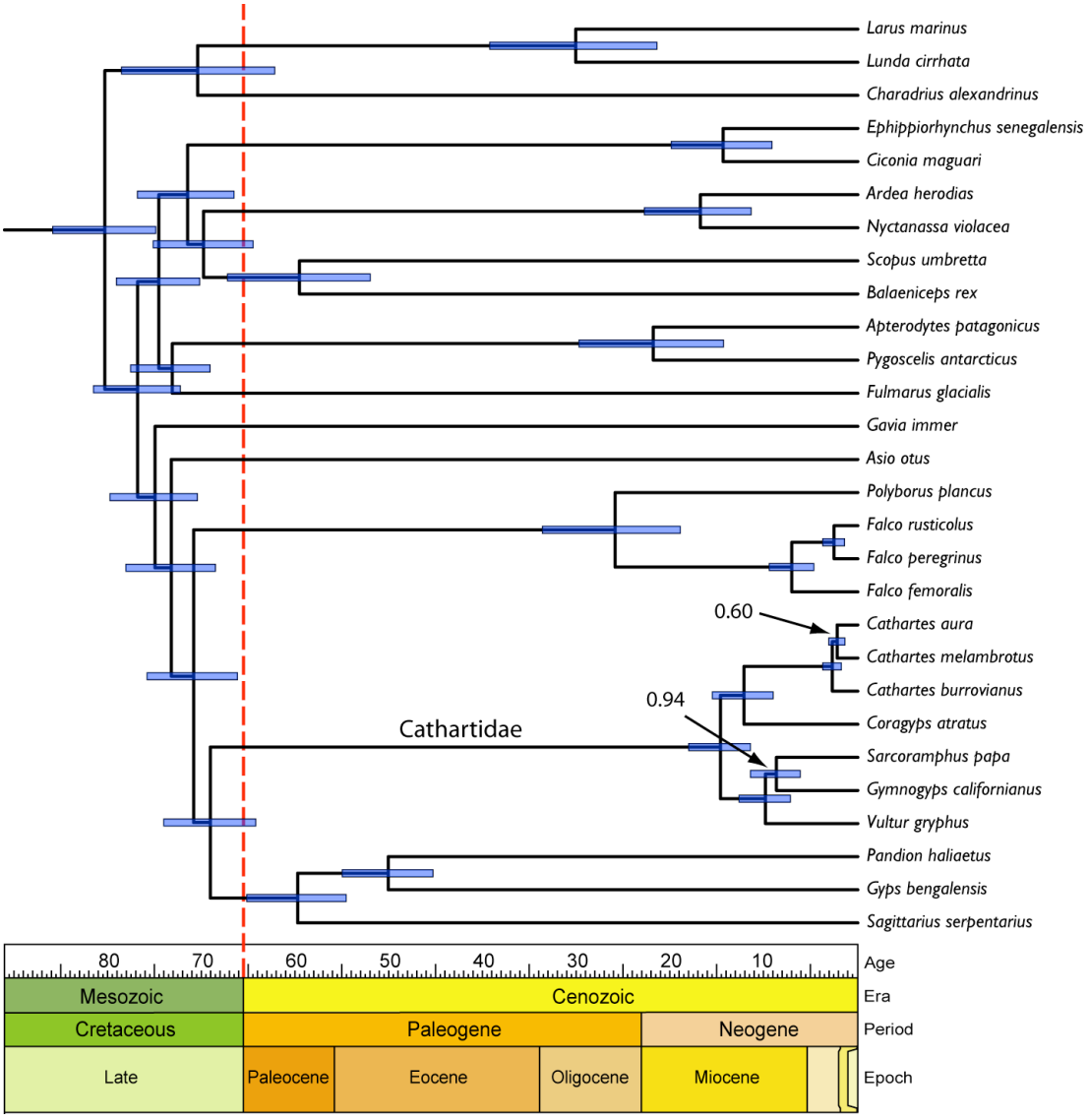

Figure 4.6 Maximum clade credibility chronogram estimated using a Bayesian modelling of rate evolution assuming a lognormal distribution of uncorrelated branch-specific substitution rates. Rate heterogeneity for each locus was described by separate but jointly-sampled lognormal distributions. The dashed vertical red line marks the K-Pg boundary, and error bars represent posterior probability (0.95) credibility intervals on inferred ages. Although Cathartidae has its origin in the Late Cretaceous, crown group cathartid taxa are quite young, diverging much later in the Miocene. Posterior node probabilities are shown only for ingroup nodes of probability < 1.0. Galloanserae outgroups not shown.

REFERENCES

Abdo, Z., V. Minin, P. Joyce, and J. Sullivan. 2005. Accounting for uncertainty in the tree topology has little effect on the decision-theoretic approach to model selection in phylogeny estimation. Molecular Biology and Evolution 22:691-703.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267-281 *in* Second International Symposium on Information Theory (P. N. Petrov, and F. Csaki, eds.). Akademiai Kiado, Budapest.

Alfaro, M. E., F. Santini, C. D. Brock, and K. Schwenk. 2007. Do reefs drive diversification in marine teleosts? Evidence from the pufferfish and their allies (Order Tetraodontiformes). Evolution 61:2104-2126.

Amadon, D. 1977. Notes on the taxonomy of vultures. Condor 79:413-416.

Avise, J. C., and W. S. Nelson. 1995. A commentary on the use of sequence data for phylogeny reconstruction - reply. Molecular Phylogenetics and Evolution 4:353-356.

Avise, J. C., W. S. Nelson, and C. G. Sibley. 1994. DNA sequence support for a close phylogenetic relationship between some storks and New World vultures. Proceedings of the National Academy of Sciences of the United States of America 91:5173-5177.

Box, G. E. P. 1976. Science and statistics. Journal of the American Statistical Association 71:791-799.

Brown, J. M., S. M. Hedtke, A. R. Lemmon, and E. M. Lemmon. 2010. When trees grow too long: investigating the causes of highly inaccurate Bayesian branch-length estimates. Systematic Biology 59:145-161.

Brown, J. W., and D. P. Mindell. 2009. Diurnal birds of prey (Falconiformes). Pages 436-439 *in* The Timetree of Life (S. B. Hedges, and S. Kumar, eds.). Oxford University Press.

Brown, J. W., R. B. Payne, and D. P. Mindell. 2007. Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson et al. Biology Letters 3:257-259.

Brown, J. W., J. S. Rest, J. García-Moreno, M. D. Sorenson, and D. P. Mindell. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biology 6.

Brown, J. W., and U. Sorhannus. in revision. A molecular genetic timescale for the evolution and diversification of the autotrophic stramenopiles (Ochrophyta): immense underestimates of putative fossil ages. PLoS ONE.

Brown, J. W., and M. van Tuinen. 2010. Evolving Perceptions on the Antiquity of the Modern Avian Tree*in* The Evolutionary History of Modern Birds (G. J. Dyke, and G. Kaiser, eds.). UC Press.

Burnham, K. P., and D. J. Anderson. 2004. Mulitmodel inference: understanding AIC and BIC in model selection. Sociological Methods Research 33:261-304.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Second edition. Springer-Verlag, New York.

Clarke, J. A., C. P. Tambussi, J. I. Noriega, G. M. Erikson, and R. A. Ketcham. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. Nature 433:305-308.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4:e88.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Ericson, P. G. P., C. L. Anderson, T. Britton, A. Elzanowski, U. S. Johansson, M. Källersjö, J. I. Ohlson, T. J. Parsons, D. Zuccon, and G. Mayr. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biology Letters 4:543-547.

Garrod, A. H. 1874. On certain muscles of birds and their value in classification. Proceedings of the Zoological Society of London 1874:339-348.

Hackett, S. J., C. S. Griffiths, G. P. Bates, and N. L. Klein. 1995. A commentary on the use of sequence data for phylogeny reconstruction. Molecular Phylogenetics and Evolution 4:350-356.

Hackett, S. J., R. T. Kimball, S. Reddy, R. C. K. Bowie, E. L. Braun, M. J. Braun, J. L. Chojnowski, W. A. Cox, K.-L. Han, J. Harshman, C. J. Huddleston, B. D. Marks, K. J. Miglia, W. S. Moore, F. H. Sheldon, D. W. Steadman, C. C. Witt, and T. Yuri.

2008. A phylogenomic study of birds reveals their evolutionary history. Science 320:1763-1768.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Harrison, C. J. O., and C. A. Walker. 1976. Birds of the British Upper Eocene. Zoological Journal of the Linnean Society 59:323-351.

Helbig, A. J., and I. Seibold. 1996. Are storks and New World vultures paraphyletic? Molecular Phylogenetics and Evolution 6:315-319.

Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Molecular Biology and Evolution 22:1561-1568.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association 90:773-795.

Kitazoe, Y., H. Kishino, P. J. Waddell, N. Nakajima, T. Okabayashi, T. Watabe, and Y. Okuhara. 2007. Robust time estimation reconciles views of the antiquity of placental mammals. PLoS ONE 2:e384.

Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. Annals of Statistics 22:79-86.

Lanave, C., G. Preparata, C. Sacone, and G. Serio. 1984. A new method for calculating evolutionary substitution rates. Journal of Molecular Evolution 20:86-93.

Ligon, J. D. 1967. Relationships of the cathartid vultures. Occasional Papers / University of Michigan, Museum of Zoology 651:1-26.

Livezey, B. C., and R. L. Zusi. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. Zoological Journal of the Linnean Society 149:1–95.

Marshall, D. C. 2010. Cryptic failure of partitioned Bayesian phylogenetic analyses: lost in the land of long trees. Systematic Biology 59:108-117.

Marshall, D. C., C. Simon, and T. R. Buckley. 2006. Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate variation and attention to branch length priors. Systematic Biology 55:993-1003.

McGuire, J. A., C. C. Witt, D. L. Altshuler, and J. V. Remsen Jr. 2007. Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. Systematic Biology 56:837-856.

Minin, V., Z. Abdo, P. Joyce, and J. Sullivan. 2003. Performance-based selection of likelihood models for phylogenetic estimation. Systematic Biology 52:674-683.

Nylander, J. A. A. 2002. MrModeltest, version 2.0. Program distributed by author.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53:47-67.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53:793-808.

Pupko, T., D. Huchon, N. Okada, and M. Hasegawa. 2002. Combining multiple data sets in a likelihood analysis: Which models are the best? Molecular Biology and Evolution 19:2294-2307.

Rambaut, A., and A. J. Drummond. 2007. Tracer, version 1.4. Available from the authors (http://tree.bio.ed.ac.uk/software/tracer/).

Rannala, B. 2002. Identifiability of parameters in MCMC Bayesian inference of phylogeny. Systematic Biology 51:754-760.

Renner, S. S., G. W. Grimm, G. M. Schneeweiss, T. F. Stuessy, and R. E. Ricklefs. 2008. Rooting and dating maples (Acer) with an uncorrelated-rates molecular clock: implications for North American/Asian disjunctions. Systematic Biology 57:795-808.

Rich, P. V. 1983. The fossil history of vultures: a world perspective. Pages 3-25 *in* Vulture Biology and Management (S. R. Wilbur, and J. A. Jackson, eds.). University of California Press, Berkeley.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Rutschmann, F. 2006. Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. Diversity and Distributions 12:35-48.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6:461-464.

Sibley, C. G., and J. E. Ahlquist. 1990. Phylogeny and Classification of Birds. Yale University Press, London.

Slack, K. E., C. M. Jones, T. Ando, G. L. Harrison, R. E. Fordyce, and D. Penny. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Molecular Biology and Evolution 23:1144-1155.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Steel, M. 2005. Should phylogenetic models be trying to 'fit an elephant'? Trends in Genetics 21:307-309.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4 (beta 10). Sinauer Associates, Sunderland, Massachusetts.

Tagliarini, M., J. Pieczarka, C. Nagamachi, J. Rissino, and E. de Oliveira. 2009. Chromosomal analysis in Cathartidae: distribution of heterochromatic blocks and rDNA, and phylogenetic considerations. Genetica 135:299-304.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research 22:4673-4680.

Wetmore, A. 1964. A revision of the American vultures of the genus Cathartes. Smithsonian Miscellaneous Collection 146:1-18.

Yang, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Molecular Biology and Evolution 10:1396-1401.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. Journal of Molecular Evolution 39:306-314.

Zhong, B., T. Yonezawa, Y. Zhong, and M. Hasegawa. 2009. Episodic evolution and adaptation of chloroplast genomes in ancestral grasses. PLoS ONE 4:e5297.

Chapter 5

All models are wrong but some are useful: identifying useful partitioned models for phylogenetic inference via posterior predictive approaches

ABSTRACT

Recent technological and methodological advances have allowed the implementation of arbitrarily complex models of molecular evolution into methods of phylogenetic reconstruction. These models have greatly increased the amount of information that can be extracted from heterogeneous data. That said, the legitimacy of all inferences made using maximum likelihood or Bayesian methodologies ultimately rely on the suitability of the underlying model assumed. While standard nucleotide substitution models are regularly tested for fit to the data at hand (e.g. using the Akaike information criterion), various partitioning schemes, which themselves represent different modellings of the molecular evolutionary process, routinely are not. Here I evaluate the fit of models of varying complexity (number of partitions) to multi-gene mitochondrial DNA alignments, with a particular interest in the treatment of edge length parameters across partitions (that is, whether they are assumed to be proportional across partitions or are left unconstrained). To arbitrate amongst models within an *a priori* candidate set of partitioned models, I extend a model selection strategy based upon a Bayesian posterior predictive approach, where trained models are ranked based upon their respective abilities to predict the composition of the empirical matrix. Ultimate model selection using this method is contrasted against established methods employing information theoretic, decision theoretic, and Bayesian approaches on empirical alignments. In general, the posterior predictive model selection approach has lower discriminating power for small matrices (in terms of the number of taxa sampled). A strong influence of branch length prior is also seen on ultimate model selection, where more stringent priors generally lead to a much poorer fit of the data, although the same is not seen using Bayes

factors. For both empirical alignments, most models that unlink branch lengths across partitions are disfavoured, although not all can be statistically rejected. Nevertheless, all methods agree that partitioning by gene is a generally very poor practice to fit molecular sequences. This should be of interest to practicing phylogeneticists, as a gene-centric view is often assumed. The method as now implemented appears to be useful for identifying poor models, but not necessarily in arbitrating amongst non-rejected models. These preliminary results suggest fertile ground for future research, in particular in constructing more sufficient summary statistics and more fully incorporating predictive ability.

INTRODUCTION

DNA technology has advanced to the point where it is relatively easy and inexpensive to generate large amounts of sequence data from multiple genomic regions. Not surprisingly, it is now becoming standard practice in phylogenetic inference to utilize data from multiple genes simultaneously. From these data it is abundantly clear that there is considerable heterogeneity in compositional makeup and evolutionary rates both within (e.g. Griffiths, 1997; Naylor and Brown, 1997; Naylor et al., 1995) and between (e.g. Cummings et al., 1995; Russo et al., 1996; Wolfe et al., 1989; Zardoya and Meyer, 1996) genes. The DNA technology revolution is thus a mixed blessing: while it has increased our understanding of biological diversity and permitted much larger sample sizes, it also brings with it the burden of developing new ways to handle these mixed data (Pupko et al., 2002).

Statistical modeling allows one to accommodate heterogeneity in the evolutionary process, and numerous models have been proposed to do this. No attempt will be made to review all of the currently available models, as several comprehensive summaries exist (Bos and Posada, 2005; Huelsenbeck et al., 2004; Liò and Goldman, 1998; Posada and Buckley, 2004; Posada and Crandall, 2001a; Swofford et al., 1996). Rather, this paper will focus on selecting from amongst these possibilities the model(s) that best captures the information in a given data set.

Model-based approaches to phylogenetic inference are today ubiquitous. This is due to myriad advantages of using statistical models, the primary one being the ability to

identify and test general principles of the underlying data (Sullivan and Joyce, 2005). Perhaps the best example of this in phylogenetic inference is the identification and modeling of among-site rate variation (Yang, 1993; Yang, 1994). An assumption of rate constancy over sites can lead to underestimated branch lengths, inaccurate molecular dating, and even incorrect trees (Sullivan and Swofford, 2001; Yang, 1996a). Clearly, statistical parameters can greatly further our understanding of evolutionary history by capturing more information in the data.

However, despite the vast advantages of using statistical models, it must be recognized that the validity of conclusions in this framework is entirely dependent on the validity of the assumed model itself. Model-based methods, such as maximum likelihood, have been shown to be fairly robust to violation of model assumptions (Sullivan and Swofford, 2001). However, this only goes so far, as maximum likelihood can be inconsistent if the assumed model is grossly insufficient (Gaut and Lewis, 1995). Furthermore, different models can have significant effects on data exploration via saturation plots (Sullivan and Joyce, 2005), branch length estimation (Buckley et al., 2001), nodal support (Buckley and Cunningham, 2002; Buckley et al., 2001; Erixon et al., 2003), tree inference (Kelsey et al., 1999; Sullivan and Swofford, 1997), and the evaluation of competing topologies (Buckley, 2002). Model selection, therefore, is a crucial component of any phylogenetic study, and should not be taken lightly.

Natural phenomena can be thought of as being governed by a very large number of (unknown) parameters whose interrelationships are very complex (Buckland et al., 1997; Sanderson and Kim, 2000). A model, by definition, is a simplification of these events. Why, then, should we not use the most complex (and, therefore, most realistic) model in our arsenal to analyze our data? While it is true that (appropriately) increasing model complexity will better approximate reality, adding parameters also carries with it two significant drawbacks. First of all, overparameterization can lead to identifiability issues (Rannala, 2002). Here, different combinations of parameter values lead to the same overall likelihood, making it impossible to determine the true values of these parameters even if the data set contains significant amounts of information. Analyses with an over-parameterized model will thus yield inaccurate and imprecise results. Rannala (2002)

shows that in a Bayesian framework such identifiability problems will increase the importance of the prior probabilities, regardless of the amount of data involved. The second, related, problem deals with whether sufficient information exists within a data set to estimate specific parameters (Buckland et al., 1997; Holder and Lewis, 2003). For example, if the ratio of data points to parameters is low, parameter estimates can be quite unreliable because of sampling errors. This is a particularly grave problem for maximum likelihood inference, where all parameters (including nuisance parameters) are jointly estimated. For Bayesian inference the problem is a little less severe as marginal likelihoods are involved, taking into account uncertainty in all parameters. That said, a low ratio of data points to parameters is never a good thing, and can never be completed accounted for through marginalization (Holder and Lewis, 2003).

It seems clear, then, that model construction should not be trying to "fit an elephant" (Steel, 2005), that is, trying to fit reality perfectly. The goal of modeling should not be to find the "true" model that generated the data, but instead to construct a model with adequate parameters to describe important features a given data set (Johnson and Omland, 2004). Furthermore, selection from a candidate set of models should be viewed as identifying the best approximating model (Buckland et al., 1997). If, for example, data are sparse in information content, they can only ever support a simple model with a small number of parameters. Interestingly, model underfitting (i.e. using too few parameters) can be a much worst problem than overfitting (Buckley and Cunningham, 2002; Erixon et al., 2003; Sullivan and Swofford, 1997), so this needs to be considered as well.

Dimensions of phylogenetic model space

Standard phylogenetic models currently in use consider molecular substitution as a homogeneous continuous-time stochastic Markov process, all sites being identical and independently distributed (i.i.d.). Such models belong to the general-time reversible (GTR) family of models (Lanave et al., 1984; Rodríguez et al., 1990; Tavaré, 1986), the most general of which takes the instantaneous rate matrix form of:

$$\mathrm{Q} = \{q_{ij}\} = \begin{pmatrix} - & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & - & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & - & f\pi_T \\ c\pi_A & e\pi_C & f\pi_C & - \end{pmatrix}$$

where a → f are substitution rate parameters (n = 6; often more concisely written as 'abcdef', listing in order the upper matrix) and π_i is the equilibrium frequency of character state i (n = 4). Although this ostensibly totals 10 substitution model parameters, the model can be simplified slightly. First, in order to prevent substitution rate parameters from becoming unidentifiable (Rannala, 2002), each is typically expressed relative to a particular parameter (often 'f') which is fixed to rate 1.0, rendering the matrix a relative-rate matrix. Second, since nucleotide frequencies must sum to 1.0, only three frequencies need be estimated. This decreases the number of GTR estimable parameters to 8.

All other GTR-family models are nested within of this most general form. Huelsenbeck *et al*. (2004) tally a total of 203 possible GTR-family models that one may implement. However, this enumeration considers only the permutation of how substitution rate parameters are set to be equal/distinct. For example, if in the above matrix the constraints $a = c = d = f$ and $b = e$ are implemented (i.e. model 'abaaba') the model collapses to the well-known transition-transversion model of Hasegawa et al. (1985):

$$\mathrm{Q} = \{q_{ij}\} = \begin{pmatrix} - & a\pi_C & b\pi_G & a\pi_T \\ a\pi_A & - & a\pi_G & b\pi_T \\ b\pi_A & a\pi_C & - & a\pi_T \\ a\pi_A & b\pi_C & a\pi_C & - \end{pmatrix}$$

Additionally, differences in the number of equilibrium state frequency parameters can be considered: zero [i.e. nucleotides are fixed to some value, possibly equal], one [four possibilities: $\pi_A/(\pi_C + \pi_G + \pi_T)$, $\pi_C/(\pi_A + \pi_G + \pi_T)$, $\pi_G/(\pi_A + \pi_C + \pi_T)$, or $\pi_T/(\pi_A + \pi_C + \pi_G)$], two [3 possibilities: $(\pi_A + \pi_C)/(\pi_G + \pi_T)$, $(\pi_A + \pi_G)/(\pi_C + \pi_T)$, or $(\pi_A + \pi_T)/(\pi_C + \pi_G)$], or three parameters [1 possibility since the fourth frequency is obtained by subtraction]; for a total of 9 ways to parameterize equilibrium states. Finally, differences in

substitution rates across sites can be accommodated through either modelling gamma-distributed relative-rate heterogeneity across sites (zero or one parameters) and/or through considering the proportion of invariable sites along the alignment (Gu et al., 1995; zero or one parameters). Considering all of these parameters, the number of possible GTR-family models increases to 7308.

There are two final components of a phylogenetic model to consider. First is the tree topology, which is typically the parameter of interest (the others being so-called 'nuisance parameters'). For T taxa, there are:

$$\frac{(2T-3)!}{2^{T-2}(T-2)!}$$

fully-bifurcating labelled rooted tree topologies. However, due to the computational convenience of Felsenstein's 'pulley principle' (Felsenstein, 1981) when using time-reversible models, most software packages instead estimate unrooted trees. Felsenstein (2004) shows that each unrooted tree with n taxa can be mapped onto one rooted tree with $T-1$ taxa. Thus, substituting $T-1$ for T above, we see the number of possible unrooted trees for T taxa is:

$$\frac{(2T-5)!}{2^{T-3}(T-3)!}.$$

In addition to tree topology, we must also consider branch lengths (BLs), which are also estimated from the data. For an unrooted tree with T taxa, there are T terminal branches and T-3 internal branches, for a total of 2T-3 BL parameters. However, in standard practice these tree-specific dimensions are typically not considered when evaluating the relative fit of competing models because these parameters are common to all models, and thus do not contribute to differences in the number/types of parameters that may distinguish models.

Partitioned models

Certainly all models entail simplifying assumptions that facilitate computation and interpretation. However, it is worthwhile to consider whether these assumptions hold, and

if they do not, what influence they may have on resulting inferences. One assumption entailed by GTR-family models that is demonstrably false is that all sites are i.i.d. (i.e. that each alignment column is statistically independent, and that each column is an identically-distributed realization of the same underlying evolutionary process). While the inclusion of gamma-distributed rate heterogeneity across sites greatly ameliorates the problem (Yang, 1996a), this involves only differences in the *relative rate* of substitution, and is of limited use if the *pattern* of substitution differs across sites. If such heterogeneity is present but not accounted for, then inference will necessarily be based upon compromised parameter estimates (i.e. parameter values will be 'averaged' across distinct processes).

A more general approach to accommodating among-site substitution heterogeneity is through the use of partitioned phylogenetic models (e.g. Nylander et al., 2004). Here, sites within a predefined subset genetic region ('partition') are assigned a substitution model that is independent of models for sites outside that partition, although all models contribute to the inference of a shared topology and (usually) shared BL parameters (but see below). The reasoning behind this approach is that, while the i.i.d. assumption may be unjustified for an alignment as whole, sites *properly organized according to some criterion* may roughly satisfy the i.i.d. assumption *within that partition*. The pertinent question then becomes *what is the optimal way to partition molecular sequence data*, or, alternatively, *how do we assess the relative fit of alternative partitioning strategies*? Importantly, we would like to be able to recognize when we might have over-partitioned our data (a form of overparameterization).

The urge to construct partitioned models by dividing data up into progressively finer categories is strong, as subsets of apparent increasingly similar characters can always be perceived. From set theory we know that the number of possible non-empty partitions of n elements is given by the appropriate Bell number,

$$B_n = \frac{1}{e}\sum_{k=0}^{\infty}\frac{k^n}{k!}.$$

Bell numbers increase very rapidly. For example, with five elements there are 52 possible partitioning strategies, and with ten elements the number of partitioning strategies increases to 115,975. While it may be possible to examine all possible partitioning strategies with small values of *n*, an investigator should guard against such 'data dredging', as this can lead to the identification of spurious effects (Burnham and Anderson, 2002). This would seem especially important when considering the dimensions of partitioned model space. Instead, the construction of models should be guided by theoretical expectations and empirical precedence.

In the present paper I strive to identify optimal partitioning strategies for mitochondrial DNA (mtDNA). The mtDNA genome has the property of not generally not undergoing recombination (Berlin and Ellegren, 2001), meaning that all sites within the genome have the exact same genealogical history. Differences amongst regions within the genome in patterns of polymorphism can thus be explained by differences in nucleotide frequencies and/or substitution rates rather than by, say, gene tree incongruence (Maddison, 1997). The partitioned models I consider here are motivated by ubiquitous empirical patterns of substitution heterogeneity across codon positions (e.g. Shapiro et al., 2006). A primary goal of this study is to determine whether these patterns are similar across genes or if they are better described on the finer scale of individual genes. I thus consider gene-specific codon positions as the fundamental units with which to construct partitioning regimes.

Each partitioned model considered here is distinguished by both the number of assumed partitions and the way in which codon positions are grouped across genes. I will use a specific notation, $M^{P}_{1,2,3}$, over the course of this paper for the identification of distinct partitioned models. Here, the superscript P represents the number of partitions within the model and the subscript identifies the three codon positions. An underlined subscript indicates that a particular codon position is linked across all genes, whereas a non-underlined subscript indicates that the relevant codon position is treated separately across genes. For example, $M^{P}_{1,2,\underline{3}}$ represents the model where 3rd codon positions are shared across P genes but 1st and 2nd positions are each treated individually within genes,

and $M^{P}_{(1+2),\underline{3}}$ represents the model where once again 3rd codon positions are shared across P genes but 1st and 2nd positions are grouped within individual genes.

Previous studies have considered amalgamating codon positions across all genes (e.g. Caterino et al., 2001; DeBry, 1999; Yang, 1996b) (my $M^{3}_{\underline{1,2,3}}$) or applying separate models for each codon position within each gene (e.g. DeBry, 1999; Reed and Sperling, 1999) (my $M^{P}_{1,2,3}$). The present study considers a more general modelling of codon positions across individual genes. For example, because of the degeneracy of the genetic code at 3rd positions, it might be reasonable to expect all 3rd positions to evolve in the same trajectory, irrespective of the gene of origin (my $M^{P}_{(1+2),\underline{3}}$ and $M^{P}_{1,2,\underline{3}}$). An alternative strategy might be to recognize the distinction of 3rd positions but group the two remaining positions, whether within genes (my $M^{P}_{1+2,3}$) or across genes (my $M^{2}_{(\underline{1+2}),\underline{3}}$); the latter model was investigated by Shapiro et al. (2006). Similar arguments can be made for the distinctiveness of highly-conserved 2nd positions within and across genes (my $M^{2}_{(\underline{1+3}),\underline{2}}$, $M^{P}_{(1+3),\underline{2}}$, $M^{P}_{(1+3),2}$, and $M^{P}_{1,\underline{2},3}$). Thus we might ask 'is my alignment better described by treating 3rd positions as excessively fast, or 2nd positions as being excessively slow?' Several potential partitioned models appear unreasonable to consider. These include any models that group conserved 2nd positions with hypervariable 3rd positions (i.e. of the form $M^{P}_{1,(2+3)}$). Additionally, given the present state of knowledge I find no justification for combining codon positions of some genes to the exclusion of other genes (e.g. 1st position of COX1 and ND2 vs. 1st position of CYTB, ND4, and ND5) or combining a codon position of one gene with a different codon position of another gene (e.g. 1st position of COX1 with 2nd position of ND2).

The treatment of branch length parameters in partitioned models

In the present paper all parameters across partitions are treated as independent except for topology, which is a linked parameter across all partitions (i.e. all partitions update the same topology parameter). This is an explicit assumption that all characters share the same evolutionary history (genealogy), which is a reasonable assumption for mtDNA. An interesting question that has received far too little attention is how to treat BL parameters in the context of partitioned model arbitration. For *P* partitions and *T* taxa there are three general possibilities: 1) assume BLs are equal across partitions (2*T*-3 parameters), 2)

assume BLs are proportional across partitions ($2T - 3 + P$ parameters), or 3) estimate separate BLs for each partition (P x ($2T$-3) parameters). Each of these models entail strong assumptions about patterns of molecular substitution. Option 1, that BLs are identical across partitions, is inappropriate for the present study (for example when comparing 2nd and 3rd codon positions), and so will not be considered here. Option 2 is an economical approach to accommodating amongst-partition heterogeneity in BL parameters, as only P extra relative-rate parameters are required. However, the assumption of strict proportionality across large T and/or large P is dubious, as no known mechanism(s) exist to predict such a result. Thus, while option 3 likely represents a more realistic model of molecular substitution, the cost of adding a large number of estimable parameters may be prohibitively high for alignments of finite length. In the present study I consider the latter two BL model options for all candidate partitioned models.

Objective of the present study

Here, I am interested in determining the optimal level of partitioning protein-coding mtDNA sequence data for phylogenetic inference. To this end I extend a criterion for use in partitioned model selection, based upon the approach of Bayesian posterior prediction. The relative efficacy of this new method is compared to existing model selection methods based upon information theory, decision theory, and Bayesian approaches through identifying optimal partitioning strategies for multi-gene protein-coding DNA alignments from paleognathous birds and primates, with specific regard to the treatment of BL parameters across partitions and how inclusion of this dimension of model space ultimately influences model selection.

STATISTICAL APPROACHES TO MODEL ARBITRATION

Given the hazards of assuming an inappropriate model, it is clear that models should not be chosen blindly (Posada and Crandall, 1998; Posada and Crandall, 2001a; Posada and Crandall, 2001b). Moreover, given the breadth of phylogenetic model space described above, it is clear that one should adopt a robust, consistent statistical approach when attempting to identify optimal model(s) for inference into a particular problem. Fortunately, a number of objective model selection criteria are currently available. As an

preamble to the novel method extended here, I briefly summarize approaches that have previously been used for phylogenetic model arbitration, commenting on their relative strengths and weaknesses (summarized in Table 5.1). All of these methods address, in different ways, the trade-off between bias (distance between estimate and truth) and variance (spread of estimates around truth; Posada and Buckley (2004)).

Likelihood ratio tests (LRT)

The likelihood ratio test (LRT) has long served as the workhorse of model arbitration in molecular phylogenetics. This is a null hypothesis approach that compares models of varying complexity in a pairwise fashion by considering both the fit of the data to the model (i.e. the likelihood) and the difference in the number of free parameters involved. Likelihood ratio tests in phylogenetic model selection are performed as follows. First, an initial starting topology is constructed, typically using a neighbour-joining algorithm (Saitou and Nei, 1987). This constitutes the topology on which all models are evaluated. Although restricting model comparisons to a single tree was seen initially as a potentially important drawback, recent studies have shown that model selection is fairly robust to differences in topology (Abdo et al., 2005; Posada and Crandall, 2001a). Second, candidate models are individually fit to the data and joint maximum likelihoods are calculated. Finally, given the respective likelihoods of two nested models, the likelihood ratio test statistic (δ) is calculated as

$$\delta = 2(\log(L_{complex}) - \log(L_{simple})),$$

where $\log(L_i)$ represents the maximized log-likelihood score for model *i*. The test statistic is then compared to a χ^2 distribution with the degrees of freedom equal to the difference between the two models in the number of free parameters estimated. Hierarchical likelihood ratio tests (hLRTs) iterate this procedure at different levels of complexity until a significant difference between models is not found.

Although significantly removing the arbitrariness of model choice in phylogenetics, likelihood ratio tests have a number of disadvantages when compared to other criteria. First, LRTs require that the models being compared are nested (i.e. that one model is a

special case of the other). This means that many candidate models simply cannot be compared to one another, greatly limiting the types of questions a researcher might ask. Second, the order of model comparisons has been found to be crucial to ultimate model choice. For example, one may start with the simplest model and add parameters (the bottom-up approach) or with the most general model and remove parameters (the top-down approach) until likelihood scores do not change significantly. Pol (2004) performed the most comprehensive analysis of this attribute of hLRTs, comparing the models selected from 32 different comparison pathways. More often than not, different pathways lead to different selected models. Third, the asymptotic χ^2 approximation used to compare models has been found to unreliable when the simpler model is equivalent to fixing a model parameter at the boundary of the parameter space of the more complex model (Posada and Buckley, 2004). Although this difficulty can potentially be overcome through either parametric bootstrapping or using a mixed a χ^2 distribution as the reference distribution, this typically is simply ignored. Fourth, LRTs compare joint maximum likelihood estimates (i.e. compare point estimates), so no error in parameter estimation is considered. Fifth, LRTs suffer from multiple testing, the probability of falsely rejecting the null hypothesis at least once in *t* tests being

$$1 - (1 - \alpha)^t,$$

where α is the significance level. Finally, the all-or-nothing null-alternative approach of the LRT gives no indication as to how much better (or worst) any model is relative to any another, only if one particular model can be outright rejected when compared to one other specific model, given an arbitrary α. Model selection, it can be argued, is not an exercise in hypothesis testing, but is instead a form of data exploration. Ideally a model selection method would allow simultaneous comparison of all candidate models, quantifying the relative fit of each.

Swofford and Sullivan (2003) describe an alternative interactive approach to likelihood ratio testing (iLRT). Rather than traversing pre-set model comparison trajectories, one starts with the most general model (GTR+I+Γ) and subtracts parameters that are most close to their fixed values in a simpler model. Comparison of the general to

the less complex model is then executed in the same fashion as the hard-coded hLRTs above. The authors show that this method often selects models that are not even examined in the standard pre-set path approaches. Despite this obvious advantage, the iLRT needlessly adds a dimension of subjectivity to model selection, as some parameters are interrelated and the choice of parameter to drop is up to the investigator. Moreover, the interactive approach does not ameliorate any of the other problems listed above. Given these drawbacks, let alone the fact that partitioned models are generally not nested, I will not consider the LRT in the present study.

Akaike information criterion (AIC)

A more appropriate approach to model selection in general is grounded in information theory. Given model i with K_i estimable parameters, Akaike (1973) developed a criterion (AIC) whereby model fit and simplicity are expressed in a common currency:

$$\mathrm{AIC_i} = -2\log(L_i) + 2K_i.$$

As with many model selection criteria, AIC is composed of two elements: a goodness of fit term (quantified through the maximized joint log-likelihood), and a term that penalizes overparameterization. AIC is an estimate of the expected Kullback-Leibler (K-L) information lost when using fitted model i to approximate the true generating model (Kullback and Leibler, 1951). Thus, the best K-L approximating model within a candidate set is that which minimizes AIC (AIC_{min}). The K-L information lost is commonly referred to as the K-L 'distance', although strictly speaking it is instead a discrepancy (i.e. a directed distance). In the expression above, the leftmost term $2K$ is a non-arbitrary bias-correction factor for the estimation of the expected K-L distance ($2K$ is often incorrectly considered an unsophisticated penalty function for adding estimable parameters to a model; (Burnham and Anderson, 2004)). Of course, we generally do not know the true generating model (and indeed, *cannot*, if one subscribes to the idea that 'reality' is infinitely dimensional); however, because we are considering K-L discrepancies, we can instead speak of *relative* directed distances, and thus the best AIC model emerges as an estimate of the best K-L model. When sample sizes are not large (say, sample size n), a small-sample bias-correction can be used:

$$\mathrm{AIC_c} = -2\log(L_i) + 2K_i\left(\frac{n}{n - K_i - 1}\right),$$

that is, the weight of the penalty term increases when sample sizes are small. In phylogenetics, the concept of *sample* size is somewhat unclear (Sullivan and Joyce, 2005). Typically n is set to the length of the alignment (since sites are assumed to be i.i.d.). A rule-of-thumb is that $\mathrm{AIC_c}$ be used over AIC whenever $n/K < 40$. However, given the ease of calculation, Burnham and Anderson (2002) recommend always using $\mathrm{AIC_c}$, for if n is large, $\mathrm{AIC_c}$ converges to AIC. Given the progressively smaller (partitioned) samples sizes dealt with here, I use $\mathrm{AIC_c}$ throughout. Unlike the LRT above, all candidate models are compared simultaneously, and no arbitrary significance level need be invoked. Instead, one can directly estimate model selection uncertainty through the calculation of Akaike weights (w_i). Given M candidate models and AIC_{min}, the AIC difference between model i and the estimated K-L best model is

$$\Delta\mathrm{AIC}_i = \mathrm{AIC}_i - \mathrm{AIC}_{min}$$

and the normalized Akaike weight of model i is

$$w_i = \frac{e^{-\frac{1}{2}\Delta AIC_i}}{\sum_{m=1}^{M} e^{-\frac{1}{2}\Delta AIC_m}}.$$

With Akaike weights in hand, one can construct a confidence set of models, and potentially generate model-averaged parameter estimates (Posada, 2008). Of course, individual Akaike weights are entirely conditional upon the constituents of the candidate pool of models, and as such must be recalculated whenever this pool changes.

Bayesian information criterion (BIC)

An popular alternative approach of penalized model selection is the Bayesian information criterion (BIC), introduced by Schwarz (1978; often referred to as the Schwarz information criterion, SIC). Unlike AIC, BIC is not based on K-L information theory. Instead, the BIC is an asymptotic approximation of the logarithm of the marginal likelihood of the model (Posada and Buckley, 2004). With n data points, BIC for model i is calculated as

$$BIC_i = -2\log(L_i) + K_i \log n .$$

Because BIC approximates the marginal log likelihood of a model, a BIC comparison of two candidate models is a rough approximation of the analogous Bayes factor (Kass and Raftery, 1995; see below). If all models are given equal prior (model) probabilities, then selecting the model with the smallest BIC is directly equivalent to choosing the model with the highest posterior probability (Posada and Buckley, 2004). However, a strict assumption of equal model priors is not necessary, as these priors can subsequently be incorporated into calculations of posterior model probabilities (sometimes referred to as BIC weights):

$$\Pr(M_i) = \frac{e^{-\frac{1}{2}\Delta BIC_i} p_i}{\sum_{m=1}^{M} e^{-\frac{1}{2}\Delta BIC_m} p_m},$$

where p_i is the model prior probability for model *i*. [In phylogenetics, where BIC is a relative newcomer to model selection, unequal model priors have yet to be investigated; that is, p_i is implicitly assumed to be $1/M$]. Regardless, as with the AIC weights above, posterior model probabilities gives one the potential to generate model-averaged inference, thereby accommodating model selection uncertainty.

Unlike all of the other model selection methods that will be discussed here, BIC is the only one that takes into explicit account the size of the data (i.e. *n* above); indeed, the magnitude of the penalty term is intimately tied to *n*, and BIC will tend to select simpler models than AIC as *n* increases. Furthermore, if the true generating model is within the candidate set (an implicit assumption made by the method), then BIC, unlike AIC, is a *consistent* model selector (i.e. it converges on the correct model as $n \rightarrow \infty$). However, the argument for consistency requires belief that 1) a 'true', non-infinitely-dimensional model exists, and 2) that the model is small enough to have found itself within the candidate set of models. If these beliefs are not met, then the argument for consistency is irrelevant (Kuha, 2004). [AIC, for example, subscribes to no such assumptions]. Regardless, all other things being equal, the relatively stronger penalty function implemented in BIC (most often leading to the selection of simpler models) tends to appeal to empirical phylogeneticists, as the allure of *parsimoniousness* is well ingrained.

However, a concern regarding BIC is that, despite being an approximation to a fully Bayesian method, prior probabilities on model *parameters* are not evident. This is problematic because the marginal likelihood of a model, the quantity which BIC attempts to approximate, is itself strongly dependent upon parameter prior probabilities. In fact, implicit priors *are* at play, just priors of a particular form. Specifically, BIC assumes the *unit information prior*, which is a multivariate normal prior centered on the joint maximized likelihood point estimate with variance equivalent to the expected information contained in a single observation (Posada and Buckley, 2004). This is a very restrictive (and idealized) assumption regarding how parameter values are expected to be distributed. In general, it is not guaranteed whether the priors an investigator might routinely attach to parameters for a particular problem would correspond at all with the priors presupposed by BIC (Weakliem, 1999). Indeed, priors typically used in Bayesian phylogenetics today are quite dissimilar from the unit information prior form. It is therefore interesting to examine the influence of parameter prior choice on ultimate model selection through comparing model choice from BIC and Bayes factors (see below); the expectation is, given the diffuse priors implicit in BIC, that BIC will tend to select simpler models than Bayes factors derived from priors implemented by an investigator.

Decision theoretic criterion (DT)

Minin *et al*. (2003) propose a method rooted in decision theory (DT) that considers *performance* when selecting from amongst a candidate set of models. The motivation of this approach is that, because searching for the 'true' model is a hopeless pursuit, we should instead focus on how alternative candidate models estimate some salient focal parameter(s) common to all models; that is, how well a particular model performs in estimating the parameter(s) relative to all other candidate models. Performance can essentially be evaluated with any set of parameters common to all models within the candidate set; for phylogenetic models, the natural choice is BL parameters as an index of how well models capture information in the data. Let B_i be the vector of estimated BLs for model i. The loss function for using model i in place of 'true' j is given by:

$$\left\| B_i - B_j \right\| = \sum_{b=1}^{2T-3} (B_{ib} - B_{jb})^2$$

that is, the squared Euclidean distance between the BL vectors estimated from the two models. This indicates the performance of a model (i.e. how well the BL estimates of model *i* agree with estimates from other models). Assuming equal prior probabilities for *M* models, model comparison in this framework is executed though the calculation of a posterior risk function:

$$R_i = \sum_{j=1}^{M} \left\| B_i - B_j \right\| \frac{e^{-\frac{1}{2}BIC_j}}{\sum_{j=1}^{M} e^{-\frac{1}{2}BIC_j}},$$

The right term in the risk function, using the BIC, is a weighting term that takes into account model fit and the number of parameters involved. Thus, BL estimation error for using model *i* compared to BLs estimated under model *j* is weighted by the posterior probability of model *j*. The optimal model from a candidate set is the one that minimizes the posterior risk. Similar to the AIC and BIC above, this approach allows for simultaneous model comparisons. As might be predicted, Minin *et al*. (2003) show that the DT-Risk method selects on average models that are simpler than those chosen by LRTs. An early concern regarding this method was the reliance on a starting topology and BLs (usually calculated heuristically from a neighbour-joining algorithm). While this does not seem to affect the other methods described above (Posada, 2001; Posada and Crandall, 2001), it appeared to be crucial in the performance-based approach. However, Abdo *et al*. (2005) recently demonstrated that varying the topology has very little effect on the inferred optimal model.

In the present study, I consider the treatment of BL parameters over partitions; when BLs are 'unlinked' across partitions these parameters are no longer common to all candidate models. Therefore, in order to apply the performance-based approach to model selection, it was necessary to first calculate weighted-average BLs across the *P* partitions (where weights are equal to the lengths *L* of the respective partitions, since BLs are expressed in units of expected numbers of substitutions per site):

$$\overline{BL_i} = \frac{\sum_{p=1}^{P} L_p \times BL_p}{L}.$$

Bayes factors

A fully Bayesian approach to model selection involves the calculation of Bayes factors (Kass and Raftery, 1995), a technique that has recently become quite popular in molecular phylogenetics (Brandley et al., 2005; Huelsenbeck et al., 2004; Nylander et al., 2004; Suchard et al., 2001). Let θ represent the suite of parameters in model M when considering data D. From Bayes' theorem we can write the posterior probability of fitted model i given D as:

$$\Pr(\theta_i \mid D, M_i) = \frac{f(D \mid \theta_i, M_i) \times f(\theta_i \mid M_i)}{\Pr(D \mid M_i)},$$

where $f(D \mid \theta_i, M_i)$ represents the likelihood, $f(\theta_i \mid M_i)$ is the prior probability density on the model parameters, and $\Pr(D \mid M_i)$ is the probability of the data given the model, common referred to as the marginal model likelihood. $\Pr(D \mid M_i)$ can be rewritten as:

$$f(D \mid M_i) = \int_{\theta_i} f(D \mid \theta_i, M_i) \times f(\theta_i \mid M_i) d\theta_i,$$

that is, it is the weighted average likelihood over all possible model parameter values, where the weighting is provided by the prior density. In the context of a single model, the marginal likelihood serves only as a normalizing factor to ensure that the posterior probability of some value of θ_i lies in the interval (0,1). It is generally not possible to calculate $\Pr(D \mid M_i)$ for phylogenetic problems with a non-trivial number of taxa (T), as this requires integrating over not only all possible substitution model parameter values, but also over all possible values for all $2T$-3 BL parameters *on each* of $(2T\text{-}5)!/2^{T\text{-}3}(T\text{-}3)!$ possible topologies (see above). Fortunately, using MCMC approximations one need never calculate this value when simply inferring a tree, as it drops out of ratios during the search. However, when comparing the relative fit of two candidate models, this is the key value of interest, as it quantifies the average likelihood of a model. Typically, the marginal model likelihood is approximated by the harmonic mean of likelihood values sampled from the stationary phase of an MCMC analysis, the approximation becoming

better with larger samples. Let M_0 and M_1 be competing models for some D. The Bayes factor is defined as:

$$BF_{1,0} = \frac{f(D \mid M_1)}{f(D \mid M_0)}.$$

The Bayes factor can thus be considered a Bayesian analogue to the LRT, except that while the latter compare maximized joint likelihoods, Bayes factors compare marginal likelihoods. Although Bayes factors compare models in a pairwise fashion similar to LRT, models need not be nested, and the order of model comparisons is arbitrary. An attractive property of the Bayes factor is that, unlike any of the model arbitration methods discussed above, uncertainty is accommodated in all model parameters, including topology. Bayes factors can be interpreted as the odds in favour of M_1 against M_0 that are given by the data, or as a comparison of the ability of the models to update the priors (Kass and Raftery, 1995). Like AIC and BIC above, Bayes factors are not evaluated according to any strict significance level. Rather, they form a continuum from low to high support. Nevertheless, a recommended scale for interpreting Bayes factors has been proposed Kass and Raftery (1995): $2\ln(BF_{1,0}) < 6$ indicates positive evidence against M_0; $6 < 2\ln(BF_{1,0}) < 10$ indicates strong evidence against M_0; and $2\ln(BF_{1,0}) > 10$ is decisive evidence against M_0 in favour of M_1.

Reversible-jump MCMC methods

A drawback of the Bayes factor method described above is that if M models are to be compared, M separate rigorous MCMC searches must first be performed to approximate the marginal likelihoods of all candidate models. However, Suchard *et al*. (2001) and Huelsenbeck *et al*. (2004) describe reversible-jump MCMC algorithms that simultaneously examine competing models of different dimensions while exploring parameter space. Analogous to conventional parameters, the amount of time an MCMC chain spends sampling a particular model is a valid approximation of the posterior probability of that model. This approach then has all of the advantages of the previous methods, including simultaneous comparisons and accounting for error in both

parameters and models. This will likely be the future of model selection for phylogenetic inference, however at this writing no software is available.

Model adequacy

Bollback (2002) argues that a selected model should not merely be the *best* amongst a candidate set, but that, trained on the empirical data, it should also be able to accurately *predict* features of the data. In a sense, model adequacy (or 'posterior predictive check') approaches are in the same spirit as the DT method above in insisting that model selection should be tied to performance rather than simply penalizing a likelihood; however, while the DT method is tied to *estimation* performance, model adequacy approaches instead focus on *observables*. I extend this method to partitioned model selection. The protocol proceeds as follows. Start by performing an MCMC phylogenetic analysis of a data set assuming a candidate partitioned model of molecular substitution. Next, take random samples from the posterior distribution of parameter combinations, including topology, branch lengths, and substitution model parameters; these values will comprise partition-specific parameter vectors. Taking these parameter vectors, simulate new data sets, the combined size of which will equal that of the empirical alignment. Finally, concatenating the sub-partitions, summarize the composition of each simulated data set using some summary statistic. Performing this for many thousand times produces a predictive distribution of the statistic. If a model is 'adequate', then the simulated and original data sets should be similar; in other words, the predictive distribution should be centered on the realized test statistic from the empirical alignment. Bollback (2002) uses the multinomial test statistic originally proposed by Goldman (1993) as a criterion for model predictiveness, and I use it here. Given an alignment of length N with p distinct site patterns (columns in the alignment matrix) of frequency n_p, the multinomial test statistic $T(X)$ summarizes the number and relative frequencies of site patterns (the 'alignment column frequency spectrum'; Figure 5.1):

$$T(X) = \left(\sum_{i}^{p} n_i \ln(n_i) \right) - N \ln(N).$$

This predictive approach can be thought of as a Bayesian analogue of the parametric

bootstrap. However, the Bayesian implementation has an enormous advantage of taking into account uncertainty in topology, branch lengths, and model parameters when comparing models, whereas the parametric bootstrap focuses solely on the maximized joint likelihood estimate (where potentially some parameters are poorly estimated because of sampling artefacts). A model can be rejected in two ways. First, the realized test statistic can lie entirely outside the predictive distribution, meaning that it has poor predictive ability. Failing that, posterior predictive *p*-values can be calculated:

$$p = \frac{1}{N}\sum_{i=1}^{N} I(T(X_i) \geq T(X_{realized})).$$

Strictly speaking, we should consider two-tailed *p*-values, as the predictive distribution may lie above or below the realized test statistic. Here I will consider a threshold value of 0.05. Perhaps the greatest boon to this approach to model selection is that it is the only one where it is possible to statistically reject all models in the candidate set (rather than settling with the best of what is available).

While the posterior predictive p-value takes care of poor models, we still require a way to rank models that are not rejected. A few approaches are possible. One can potentially select the model which generates the posterior predictive p-value closest to 0.5, however this seems less than ideal given the amount of information ignored (Figure 5.2). Alternatively, one may select the most parsimonious model (i.e. that with the fewest number of parameters) from the set of models which can not be rejected. However, this too is also problematic when comparing models of the same dimensionality without appealing to some alternative criterion. A fruitful approach may be to accommodate predictive variance into how we rank models, that is, all else being equal, models are penalized by the variance (breadth) of the posterior predictive distribution of T(X). Such an approach is described by Laud and Ibrahim (1995), where the criterion of interest is the variance-penalized 'distance' between the realized test statistic and the mean predictive value:

$$D_{LI}^{\ 2} = (E[T(X)] - T(X_{realized})) + \mathrm{var}[T(X)].$$

Here I will present the square root of this value, D_{LI}, as this is more readily interpreted as a distance along the T(X) axis.

Ultimate success of the posterior predictive approach requires a test statistic (or series of test statistics) that is sufficient to discriminate the predictive abilities of alternative candidate models. Consideration of the multinomial test statistic reveals that its discriminating power may be low: treating alignment columns as unlabelled strings means that alignments of potentially disparate column compositions can nevertheless give rise to similar test statistics; indeed, it is possible for several alignments that do not share a single column to generate the exact same test statistic (Figure 5.2). In this particular case, we would not consider the approach as identifying models that can *predict the empirical alignment*, but instead models that *generate a similar level of heterogeneity in relative pattern counts*. The objective of this paper is to evaluate the performance of this criterion, especially in respect to more established model selection criteria.

Empirical mtDNA data

To explore the influence of data partitioning on empirical phylogenetic inference, I study two exemplar mtDNA matrices. The first is a 3 gene alignment of 4 primate taxa (*Homo*, *Pan*, *Gorilla*, and *Pongo*). This matrix is compelling for two reasons. First, it has proven historically to be a challenging problem to solve, however we can regard the tree now as known. Second, the small size of the matrix allows for particularly thorough analysis. The second empirical example is a 5 gene alignment paleognathous birds, a clade which is composed of the flightless Ratites (ostrich, rhea, emu, cassowary, kiwi, and the extinct moa and elephant birds) and the volant tinamous (47 extant species). The present study includes representatives from all 6 major extant lineages, 2 extinct moas, and two neognathous outgroup taxa. Ratites, being large and flightless, are a group that has traditionally been used as an example of how plate tectonics can explain evolutionary and (Gondwanan) biogeographic patterns. However, exact relationship amongst ratite taxa had proven controversial even with the mtDNA analyzed here (Cooper et al., 2001; Haddrath and Baker, 2001). Recently the monophyletic status of ratites has been

contested by both nuclear (Harshman et al., 2008) and mtDNA (Phillips et al., 2010) data, further obfuscating the role of plate tectonics in the biogeography of this clade.

All sequences were obtained from complete mtDNA genome sequences GenBank (Table 5.2). Analyses are limited to the largest protein coding genes (COX1, CYTB, ND2, ND4, and ND5) in the mtDNA genome. This is done for three reasons. First, it is clear that increasing the number of partitions necessarily decreases the number of data points per partition. As sampling error increases inversely with sample size, only the largest genes available are used so that parameter estimates would be reasonably precise. Second, the number of possible partitioning strategies increases rapidly with the number partitions (see above). To keep the number of permutations manageable I limited the study to five genes. Finally, partitioning the entire mitochondrial genome (e.g. most generally by gene and codon position) would require an exorbitant number of parameters to be estimated (e.g. if each partition is best fit by a GTR+I+Γ model of substitution, this would require 13 x 3 x 10 = 390 substitution model parameters and 13 x 23 = 897 BL parameters, for a total of 1287 free parameters).

Each gene was separately aligned using CLUSTALW v1.4 (Thompson et al., 1994) as implemented in BIOEDIT v7.05 (Hall, 1999) after first translating the nucleotide sequences into amino acids. Unalignable regions (rare, and usually at the 5' end of genes) were identified by eye and excluded from all subsequent analyses. Following alignment, genes were reverse-translated back into nucleotide sequences for use in phylogenetic reconstructions. The complete alignment amounted to 4491 bp (Table 5.3) for the primate data set and 6936 bp for the paleognaths (Table 5.4). For each separate potential process partition, likelihood scores for various substitution models were calculated in PAUP* version 4.0 (Swofford, 2003). Optimal models were selecting using AIC as calculated in MrModelTest 2.2 (Nylander, 2004).

Phylogenetic reconstruction

Multi-partition Bayesian phylogenetic analyses were carried out in MRBAYES v3.1.2 (Ronquist and Huelsenbeck, 2003). Partition-specific substitution models were identified using AIC. For each analysis, 4 Markov chains (3 hot, one cold) were run for 2.5×10^7

generations, as this delivered consistent results in preliminary analyses. To accommodate possible autocorrelation of parameters, chains were thinned so that sampling occurred at every 10^3 generations (yielding 2.5 x 10^4 samples from the posterior distribution). Each analysis was run four times from random positions in parameter space to check for convergence of the Markov chains. Convergence and effective sample sizes (ESS) were inspected through viewing generation plots in the program TRACER v1.5 (Rambaut and Drummond, 2003). Burnin was very conservatively defined as the first 5 x 10^6 generations (i.e. most analyses converged well before this point). Parameter estimates were gleaned from the combined post-burnin samples from all replicate runs. The most parameter-rich model for the paleognath matrix ($M^{15}_{1,2,3}$) was also run four times for 5 x 10^7 generations to investigate convergence and mixing. Maximum likelihood phylogenetic reconstruction was performed in RAxML v7.2.7 (Stamatakis, 2006). All partitions were assigned GTR+G substitution models, as RAxML does not support simpler models.

A computer program

To accomplish the posterior predictive calculations above a computer program was written in C++ which currently runs on a Unix platform. The program takes as input the MrBayes Nexus input file and all log (parameter and tree) files from an MCMC analysis (or multiple analyses). The files are automatically parsed to determine which models and partitioning scheme were implemented by the user. The user can determine the burnin (the number of samples to disregard from the beginning of the log files), thinning (taking every *n* samples), whether all remaining samples are used or if random samples are desired, and how many replicate simulated alignments are to be generated per posterior parameter vector (i.e. since these are based on stochastic models, one predictive alignment is not necessarily representative of the relevant parameter vector; simulating multiple predictive alignments per parameter vector accommodates this uncertainty as well). Sequence generation is accomplished using C code from seq-gen (Rambaut and Grassly, 1997). This software will eventually be made open source and released to the community. Because of the computational demands of an alternative model selection procedure based upon minimizing posterior predictive risk (Gelfand and Ghosh, 1998; to

be presented elsewhere), matrix size factors greatly into the time required for analysis, and thus the number of replicate simulated matrices that can be tractably generated. Therefore, because the primate matrix is small, I took 25000 random samples from the posterior distribution, and for each simulated 10 posterior predictive alignments. Because the paleognath data set contains more taxa, only 1 simulated predictive alignment for each of 25000 random samples were generated. I note that running model adequacy analyses alone run quite quickly, even for large number of taxa. This option has just been added to the program.

RESULTS AND DISCUSSION

Empirical model selection: a comparison to other criteria

Unlike with Bayes factors, we see a strong influence of branch length prior on the ultimately selected partitioned model using the posterior predictive approach, although this differs across alignments. Posterior predictive model adequacy for the primate data set using default MrBayes BL-priors ($\lambda = 10$) gives distressing results (Figure 5.3): not one model can be rejected as inadequate, despite the fact that the other criteria discriminate strongly across models (Table 5.5, 5.6). However, as BL-prior gets narrower, more and more models begin to get rejected (Figure 5.4,5.5). Although marginal model likelihoods decrease with increasing λ (Table 5.6), these are not of a magnitude to generate decisive Bayes factors (Kass and Raftery, 1995).

The optimally-selected model according to D_{LI} differs across BL-priors. For the λ=10 analyses, a six partition model is selected ($M^6_{1+2,3}$; where gene-specific 1^{st}+2^{nd} codon positions are partitioned to the exclusion of gene-specific 3^{rd} codon positions) with linked BL parameters across partitions. For the λ=20 analyses, the optimal model ($M^3_{\underline{1},\underline{2},\underline{3}}$; partitioned by grouped codon positions) was one in which BLs were unlinked, although the analogous linked-BL model fared nearly as well. Finally, for the most stringent BL-prior (λ=100) the optimal model was again $M^3_{\underline{1},\underline{2},\underline{3}}$, but with BLs linked across partitions. Taking the minimized D_{LI} across all models and BL-priors, we see that the best model is $M^3_{\underline{1},\underline{2},\underline{3}}$ with BLs unlinked across partitions with the exp(20) BL-prior (with D_{LI} =

145.08), although the $M^6_{1+2,3}$ with exp(10) BL-priors was a close second (with D_{LI} = 146.37).

The established model selection methods show some consensus on optimally identified partitioning strategy, and all selected models with zero uncertainty (i.e. $w \sim 1.0$, and very large Bayes factors; only marginal log likelihoods are presented). AIC, AIC_c (Table 5.5), and Bayes factors (Table 5.6) choose the most general partitioning scheme ($M^9_{1,2,3}$; partitioned by gene and codon position) with linked BL parameters across partitions. BIC and DT methods, however, prefer a much simpler three partition model ($M^3_{1,2,3}$; partitioned by codon only), also with linked BLs; this is the same model as preferred for the model adequacy approach under exp(100) BL-priors. This speaks to the stronger penalty for overparameterization in the BIC (the BIC also being a component of the DT method implemented here). It is interesting that BIC and BF disagree so strongly on the optimal model, as the former was devised as an approximation to the latter. Regardless, all conventional approaches to model selection strongly disfavour the unlinked-BL models.

In general, the model adequacy approach displayed greater discriminating power for the paleognath data set (i.e. rejecting more model according to posterior predictive p-values at the 0.05 threshold), perhaps due to the deeper tree and/or larger taxon sampling (Table 5.7; Figures 5.6-5.8). As with the primate example above, this criterion selected a novel partitioned model for relaxed BL-priors ($M^{11}_{1,2,3}$; only 2nd codon positions grouped across genes), an unlinked-BL model for exp(20) BL-priors (again $M^{11}_{1,2,3}$), and still a third model for the most stringent BL-priors ($M^{15}_{1,2,3}$) with BLs linked across partitions. For this latter prior, all but two unlinked-BL models are rejected.

The models selected by the conventional criteria for the primate alignment were identical to those selected for the paleognath alignment (Tables 5.7, 5.8). This is interesting, as Holder et al. (2008) recently argued that partitioned models are more used for longer (older) trees, while shorter trees are better described by unpartitioned models. The results presented here, based on two very different trees (both in terms of taxa and age) nevertheless converge on the same partitioned models within a given criterion, suggesting that a general (partitioned) model of mtDNA sequences may be of use. The

harmonic mean log model likelihood is much more influenced by BL-prior for the paleognath data set (Table 5.8), likely being a function of the size of the data.

Four points emerge from the results generated here. First, mtDNA sequences do not appear to be best explained by unlinked-BL models, regardless of the model selection criterion subscribed to. This is surprising, as there is no known mechanism to invoke to explain the phenomenon. Rather, it may just be a matter of power, and that departures from proportional branch lengths across partitions will be observed with larger (more taxon) matrices. However, these same results were recently found for a 42 taxon mtDNA study of cathartid vultures and outgroups (Brown et al., to be submitted-a; Chapter 4) (but interestingly not when considering mtDNA vs. nuclear DNA). Generality of the result, if such a generality exists, may be limited to birds and/or mtDNA, as the assumption of proportional BLs is rejected elsewhere (Brown et al., to be submitted-b). The results presented here are specific to mitochondrial DNA where every nucleotide is physically linked, such that all sites shared the exact same tree topology, and thus potential heterogeneity in BL parameters is constrained. However, these findings will not likely be relevant to nuclear loci that experience recombination between (or within) genes so that different loci can potentially have different histories (Maddison, 1997).

Second, mtDNA also appears to be poorly described using models which partition by gene (my M^{P}_{gene} models), regardless of the model selection criterion employed. This is particularly seen using the posterior predictive D_{LI} distance, where gene-specific partitioned models are judged as worst than unpartitioned models (M^{1}) for the paleognath alignment (and nearly so for the primate alignment). This poor fitting should be of interest to empirical phylogeneticists, as genes are typically treated as independent. Third, the posterior predictive approach appears to be much more sensitive to BL-prior than conventional Bayesian model selection based upon marginal model likelihoods. This sensitivity may be viewed as beneficial, as the approach may then be used to guide our construction of priors, and not just BL-priors but potentially priors on several different parameters in our phylogenetic models.

Finally, while the posterior predictive method appears to work to identify poor models, identifying optimal models from those that are not rejected is less

straightforward. The suitability of the D_{LI} distance is in question, as it equally weights mean predictive error with predictive variance. In practice, we may be more interested in one quantity than the other, and hence might like to weight the terms differently. However, a more severe evaluation of the approach as implemented here is that a great deal of information is lost when boiling down an alignment to a single summary statistic, and in particular with the multinomial test statistic which ultimately deals with unlabelled alignment columns (Figure 5.2). While the construction of additional alternative test statistics may facilitate the exclusion of further models, the predictive ability of a model tied to the observed empirical alignment (and not just a summary statistic) appears to provide a superior way to proceed (see below).

CONCLUSIONS AND FUTURE WORK

Posterior predictive approaches have numerous attractive properties. Philosophically, model 'ability' is tied to predicting observables, rather than penalizing likelihood (with assumptions of how error is distributed) or estimation error. That said, implementing the approach is problematic. Model adequacy approaches using solely the multinomial statistic ignore a lot of information, such that the discriminating power of the method is poor. A fruitful next step would seem to be to focus predictive ability only on those sites actually observed in the empirical alignment; such an approach is available (Gelfand and Ghosh, 1998) and is in development. Finally, ideally many simulated matrices under known conditions should be constructed to explore the sensitivity of this and other model selection criteria to such parameters as tree depth, taxon sampling, and tree balance.

ACKNOWLEDGMENTS

Paul Lewis for discussion regarding posterior prediction. I thank Derrick Zwickl, David Swofford, and Michael Sanderson for useful advice regarding preliminary presentations of this work. Elliott Sober for philosophical discussions on the methodological foundations of alternative model selection strategies. Paul Joyce for discussion regarding BIC. Jonathan Bollback for advice regarding posterior predictive distributions. Part of this work (specifically, replicate MrBayes analyses) was carried out by using 1) the resources of the Computational Biology Service Unit from Cornell University (partially

funded by Microsoft Corporation), and 2) the Center for Comparative Genomics PhyloCluster at the California Academy of Sciences. I thank Brian Simison at the California Academy of Sciences for granting me the privilege of acting as a beta tester on the PhyloCluster. Funding was provided by the University of Michigan Museum of Zoology (UMMZ), and a Rackham Predoctoral Fellowship from the University of Michigan Rackham Graduate School.

Table 5.1 Available statistical approaches to phylogenetic model selection.

	hLRT	iLRT	Parametric bootstrap	AIC	BIC	Averaging[1]	DT-Risk	Posterior prediction	Reversible-jump MCMC	Bayes factors
Model(s) selected	Best	Best	Best	Best	Best	Best set	Performance	Adequate	Best set	Best
Must models be nested	Yes	Yes	No	No	No	No	No	No	No	No
Single topology considered	Yes	Yes	Yes	Yes	Yes	Yes	Yes[2]	No	No	No
Incorporate sample size	No	No	No	No	Yes	No	Yes	No	No	No
Simultaneous comparisons	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes[3]
Account for parameter error	No	No	No	No	No	Yes[4]	No	Yes	Yes	Yes
Account for model error	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Explicitly penalize parameters	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Computational expense	Low	Moderate	High	Low	Low	Low	Low	Very high	High	Very high
Automated software	Yes	No	No	Yes	Yes	Yes	Yes	Yes[5]	Yes	No
Simultaneous model selection and inference	No	No	No	No	No	No	No	No	Yes	No

[1] Can be used with AIC or BIC using model weights
[2] Software currently available requires a single topology. However, Abdo et al. (2005) recently extended the method to accommodate tree uncertainty.
[3] Comparisons are pairwise, but no sequence of comparisons is involved.
[4] Only in the sense that the different models being combined are each allowed to have their own parameter values. Within a model, parameter error is not accommodated.
[5] While software exists to generate predictive data sets and calculate summary statistics, the original MCMC analyses for all candidate models must be performed manually.

Table 5.2 Taxa involved in this study.

Scientific name	Common name	GenBank accession number
Gorilla gorilla	Gorilla	D38114
Homo sapiens	Human	J01415
Pan troglodytes	Chimpanzee	D38113
Pongo pygmaeus	Orangutan	D38115
Tinamus major	Great Tinamou	NC_002781
Eudromia elegans	Elegant Crested Tinamou	NC_002772
Dinornis giganteus	Giant Moa	NC_002672
Emeus crassus	Eastern Moa	NC_002673
Anomalopteryx didiformis	Little Bush Moa	NC_002779
Apteryx haastii	Great Spotted Kiwi	NC_002782
Dromaius novaehollandiae	Emu	NC_002784
Casuarius casuarius	Southern Cassowary	NC_002778
Rhea americana	Greater Rhea	NC_000846
Pterocnemia pennata	Lesser Rhea	NC_002783
Struthio camelus	Ostrich	NC_002785
Anseranas semipalmata	Magpie Goose	NC_005933
Gallus gallus	Chicken	NC_001323

Table 5.3 Characteristics of genes involved in the primate study. For all partitions unalignable nucleotides have been excluded.

Gene	Partition	Characters	Polymorphic	Parsimony	AIC model
COX1	All	1539	338 (22.0%)	71 (4.6%)	GTR+Γ
	$1^{st}+2^{nd}$	1026	52 (5.1%)	12 (1.2%)	HKY+I
	$1^{st}+3^{rd}$	1026	328 (32.0%)	70 (6.8%)	GTR+Γ
	1^{st}	513	42 (8.2%)	11 (2.1%)	SYM+I
	2^{nd}	513	10 (1.9%)	1 (0.0%)	HKY
	3^{rd}	513	286 (55.7%)	59 (11.5%)	GTR+I
CYTB	All	1140	266 (23.3%)	61 (5.4%)	HKY+Γ
	$1^{st}+2^{nd}$	760	85 (11.2%)	23 (3.0%)	GTR+I
	$1^{st}+3^{rd}$	760	241 (31.7%)	56 (7.4%)	GTR+Γ
	1^{st}	380	60 (15.8%)	18 (4.7%)	GTR+I
	2^{nd}	380	25 (66.%)	5 (1.3%)	HKY+I
	3^{rd}	380	181 (47.6%)	38 (10%)	HKY+I+G
ND5	All	1812	490 (27.0%)	102 (5.6%)	HKY+I
	$1^{st}+2^{nd}$	1208	185 (15.3%)	31 (2.6%)	GTR+I
	$1^{st}+3^{rd}$	1208	425 (35.2%)	90 (7.4%)	GTR+Γ
	1^{st}	604	120 (20.0%)	19 (3.1%)	GTR+I
	2^{nd}	604	65 (10.8%)	12 (2.0%)	HKY+I
	3^{rd}	604	305 (50.5%)	71 (11.7%)	GTR+I
All	All	4491	1097 (24.4%)	234 (5.2%)	GTR+Γ
	$1^{st}+2^{nd}$	2994	322 (10.7%)	66 (2.2%)	GTR+I
	$1^{st}+3^{rd}$	2994	994 (33.2%)	216 (7.2%)	GTR+I
	1^{st}	1497	222 (14.8%)	48 (3.2%)	GTR+Γ
	2^{nd}	1497	100 (6.7%)	18 (1.2%)	HKY+I
	3^{rd}	1497	772 (51.6%)	168 (11.2%)	GTR+I

Table 5.4 Characteristics of genes involved in the paleognath study. For all partitions unalignable nucleotides have been excluded. Outgroup taxa are included in these calculations

Gene	Partition	Characters	Polymorphic	Parsimony	AIC model
COX1	All	1554	563 (36.2%)	431 (27.7%)	GTR+I+Γ
	$1^{st}+2^{nd}$	1036	90 (8.69%)	41 (3.9%)	GTR+I+Γ
	$1^{st}+3^{rd}$	1036	549 (53.0%)	425 (41.0%)	GTR+I+Γ
	1^{st}	518	76 (14.7%)	35 (6.7%)	GTR+I+Γ
	2^{nd}	518	14 (2.7%)	6 (1.2%)	HKY+I
	3^{rd}	518	473 (91.3%)	390 (75.3%)	GTR+I+Γ
CYTB	All	1146	467 (40.8%)	330 (28.8%)	GTR+I+Γ
	$1^{st}+2^{nd}$	764	132 (17.3%)	70 (9.2%)	GTR+I+Γ
	$1^{st}+3^{rd}$	764	444 (58.1%)	310 (40.6%)	HKY+I+Γ
	1^{st}	382	99 (25.9%)	50 (13.1%)	SYM+I+Γ
	2^{nd}	382	33 (8.6%)	20 (5.2%)	HKY+I
	3^{rd}	382	335 (87.7%)	260 (68.1%)	GTR+I+Γ
ND2	All	1041	558 (53.6%)	414 (39.8%)	GTR+I+Γ
	$1^{st}+2^{nd}$	694	237 (34.1%)	(62.9%)	GTR+I+Γ
	$1^{st}+3^{rd}$	694	490 (70.6%)	375 (54.0%)	GTR+I+Γ
	1^{st}	347	169 (48.7%)	110 (31.7%)	GTR+I+Γ
	2^{nd}	347	68 (19.6%)	39 (11.2%)	GTR+I+Γ
	3^{rd}	347	321 (92.5%)	265 (76.4%)	GTR+I+Γ
ND4	All	1383	688 (49.7%)	511 (36.9%)	GTR+I+Γ
	$1^{st}+2^{nd}$	922	266 (28.9%)	167 (18.1%)	GTR+I+Γ
	$1^{st}+3^{rd}$	922	605 (65.6%)	466 (50.5%)	GTR+I+Γ
	1^{st}	461	183 (39.7%)	122 (26.5%)	GTR+I+Γ
	2^{nd}	461	83 (18.0%)	45 (9.8%)	GTR+I+Γ
	3^{rd}	461	422 (91.5%)	344 (81.5%)	GTR+I+Γ
ND5	All	1812	957 (52.8%)	705 (38.9%)	GTR+I+Γ
	$1^{st}+2^{nd}$	1208	381 (31.5%)	235 (19.4%)	GTR+I+Γ
	$1^{st}+3^{rd}$	1208	820 (67.9%)	621 (51.4%)	GTR+I+Γ
	1^{st}	604	250 (41.3%)	154 (25.5%)	GTR+I+Γ
	2^{nd}	604	134 (22.1%)	83 (13.7%)	GTR+I+Γ
	3^{rd}	604	573 (94.6%)	468 (77.5%)	GTR+Γ
All	All	6936	3233 (46.6%)	2391 (34.5%)	GTR+I+Γ
	$1^{st}+2^{nd}$	4624	1109 (24.0%)	664 (14.3%)	GTR+I+Γ
	$1^{st}+3^{rd}$	4624	2898 (62.7%)	2197 (47.5%)	GTR+I+Γ
	1^{st}	2312	777 (33.6%)	471 (20.4%)	GTR+I+Γ
	2^{nd}	2312	332 (14.4%)	193 (8.3%)	GTR+I+Γ
	3^{rd}	2312	2124 (91.9%)	1727 (74.7%)	GTR+I+Γ

Table 5.5 Identifying the optimal partitioning strategy for the primate data alignment based on the maximized joint log-likelihood. Bold values indicate optimal models per criterion. L = linked-BL; UL = unlinked-BL.

Partitioned Models (L/UL)	-ln(*L*)	Parameters (n_{BL})	AIC	*w*	AIC_c	w_c	BIC	w_{BIC}	DT-Risk
M^1	11171.98	14 (5)	22372	0	22372	0	22462	0	0.10135
$M^2_{1+2,3}$ (L)	10337.32	23 (5)	20721	0	20721	0	20868	0	0.05389
$M^2_{1+3,2}$ (L)	10735.35	23 (5)	21517	0	21517	0	21664	0	1.24782
$M^3_{1,2,3}$ (L)	10219.52	32 (5)	20503	0	20504	0	**20708**	**1**	**0**
M^3_{gene} (L)	11133.58	32 (5)	22331	0	22332	0	22536	0	1.89197
$M^4_{1+2,3}$ (L)	10704.86	41 (5)	21492	0	21492	0	21755	0	2.15146
$M^4_{1+3,2}$ (L)	10276.51	41 (5)	20635	0	20636	0	20898	0	0.02782
$M^6_{1+2,3}$ (L)	10255.31	59 (5)	20629	0	20630	0	21007	0	0.10803
$M^6_{1+3,2}$ (L)	10673.25	59 (5)	21464	0	21466	0	21843	0	2.36053
$M^7_{1,2,3}$ (L)	10166.86	68 (5)	20470	0	20472	0	20906	0	0.17688
$M^7_{1,2,3}$ (L)	10155.43	68 (5)	20447	0	20449	0	20883	0	0.16003
$M^7_{1,2,3}$ (L)	10145.04	68 (5)	20426	0	20428	0	20862	0	0.03086
$M^9_{1,2,3}$ (L)	10123.89	86 (5)	**20420**	**1**	**20423**	**1**	20971	0	0.19078
$M^2_{1+2,3}$ (UL)	10331.87	28 (10)	20720	0	20720	0	20899	0	0.30425
$M^2_{1+3,2}$ (UL)	10731.75	28 (10)	21520	0	21520	0	21699	0	2.38310
$M^3_{1,2,3}$ (UL)	10214.49	42 (15)	20513	0	20514	0	20782	0	0.31601
M^3_{gene} (UL)	11130.00	42 (15)	22344	0	22345	0	22613	0	5.54784
$M^4_{1+2,3}$ (UL)	10698.97	56 (20)	21510	0	21511	0	21869	0	5.09490
$M^4_{1+3,2}$ (UL)	10262.97	56 (20)	20638	0	20639	0	20997	0	0.30605
$M^6_{1+2,3}$ (UL)	10238.86	84 (30)	20646	0	20649	0	21184	0	0.91594
$M^6_{1+3,2}$ (UL)	10663.55	84 (30)	21495	0	21498	0	22034	0	5.09617
$M^7_{1,2,3}$ (UL)	10155.03	98 (35)	20506	0	20510	0	21134	0	0.90127
$M^7_{1,2,3}$ (UL)	10138.32	98 (35)	20473	0	20477	0	21101	0	0.92596
$M^7_{1,2,3}$ (UL)	10127.03	98 (35)	20450	0	20454	0	21078	0	0.31408
$M^9_{1,2,3}$ (UL)	10103.02	126 (45)	20458	0	20465	0	21266	0	0.88477

Table 5.6 Identifying optimal partitioned-models for the primate alignment using marginal model likelihoods. BL priors are denoted as BL-X, where X is the exponential rate parameter. H.M. –ln(*L*) is the estimated harmonic mean model log likelihood. Pr(T(X)) is the posterior predictive p-value of the realized empirical multinomial test statistic. D_{LI} is the variance penalized distance between the realized empirical test statistic and the mean of the predictive distribution of the statistic. Values in bold represent the optimal partitioned-model for a given alignment/BL-prior determined from Bayes factors and D_{LI}. * indicates that the model is rejected at the $\alpha = 0.05$ level (two-sided). L = linked-BL; UL = unlinked-BL.

	Parameters	**BL-10**			**BL-20**			**BL-100**		
Model	**(n_{BL})**	**H.M. –ln(L)**	**Pr(T(X))**	**D_{LI}**	**H.M. –ln(L)**	**Pr(T(X))**	**D_{LI}**	**H.M. –ln(L)**	**Pr(T(X))**	**D_{LI}**
M^{1}	14 (5)	11179.57	0.323	179.50	11179.27	0.373	173.40	11189.65	0.614	165.63
$M^{2}_{1+2,3}$ (L)	23 (5)	10348.66	0.366	154.02	10350.66	0.405	150.88	10360.32	0.671	159.33
$M^{2}_{1+3,2}$ (L)	19 (5)	10767.46	0.343	167.69	10767.35	0.380	160.98	10774.36	0.635	162.57
$M^{3}_{1,2,3}$ (L)	28 (5)	10238.31	0.392	150.41	10241.23	0.431	146.75	10249.06	0.655	**155.02**
M^{3}_{gene} (L)	24 (5)	11149.76	0.886	236.01	11152.76	0.912	254.27	11172.62	0.980*	347.37
$M^{4}_{1+2,3}$ (L)	37 (5)	10743.69	0.331	158.19	10745.58	0.368	153.83	10773.16	0.646	155.11
$M^{4}_{1+3,2}$ (L)	38 (5)	10300.93	0.347	167.21	10300.08	0.394	161.52	10312.31	0.829	211.61
$M^{6}_{1+2,3}$ (L)	52 (5)	10291.98	0.563	**146.37**	10296.36	0.606	149.40	10304.28	0.842	203.72
$M^{6}_{1+3,2}$ (L)	46 (5)	10743.77	0.591	158.20	10729.53	0.374	163.32	10753.22	0.814	205.22
$M^{7}_{1,2,3}$ (L)	52 (5)	10220.28	0.612	149.22	10217.21	0.658	154.60	10224.64	0.870	216.33
$M^{7}_{1,2,3}$ (L)	58 (5)	10208.32	0.596	147.74	10207.96	0.643	153.94	10216.18	0.863	213.36
$M^{7}_{1,2,3}$ (L)	52 (5)	10194.59	0.352	155.05	10194.24	0.388	151.70	10209.15	0.661	157.12
$M^{9}_{1,2,3}$ (L)	67 (5)	**10181.70**	0.582	146.68	**10184.20**	0.626	151.88	**10197.27**	0.854	209.17
$M^{2}_{1+2,3}$ (UL)	28 (10)	10356.67	0.380	153.04	10356.03	0.473	146.90	10385.89	0.924	255.03
$M^{2}_{1+3,2}$ (UL)	24 (10)	10770.41	0.325	169.28	10769.57	0.384	161.19	10781.09	0.755	185.42
$M^{3}_{1,2,3}$ (UL)	38 (15)	10243.29	0.386	150.55	10249.22	0.495	**145.08**	10281.81	0.950	277.71
M^{3}_{gene} (UL)	34 (15)	11155.16	0.910	251.87	11161.33	0.948	288.93	11194.78	0.999*	499.87
$M^{4}_{1+2,3}$ (UL)	52 (20)	10757.08	0.297	165.97	10764.39	0.417	148.98	10818.76	0.963	295.57
$M^{4}_{1+3,2}$ (UL)	53 (20)	10299.99	0.391	161.71	10302.96	0.528	155.42	10338.43	0.996*	438.67
$M^{6}_{1+2,3}$ (UL)	77 (30)	10301.93	0.631	152.79	10311.62	0.829	200.18	10388.92	1.000*	672.98
$M^{6}_{1+3,2}$ (UL)	71 (30)	10736.64	0.328	171.65	10745.54	0.466	155.44	10798.94	0.998*	447.09
$M^{7}_{1,2,3}$ (UL)	82 (35)	10228.27	0.651	154.55	10235.43	0.845	205.29	10324.40	1.000*	688.48
$M^{7}_{1,2,3}$ (UL)	88 (35)	10216.17	0.645	153.99	10224.84	0.858	210.99	10310.15	1.000*	740.13
$M^{7}_{1,2,3}$ (UL)	82 (35)	10196.44	0.249	176.08	10201.98	0.399	150.37	10237.31	0.988*	352.02
$M^{9}_{1,2,3}$ (UL)	107 (45)	10193.78	0.573	146.92	10206.42	0.823	196.65	10289.17	1.000*	752.00

Table 5.7 Identifying the optimal partitioning strategy for the paleognath data alignment based on the maximized joint log-likelihood. Bold values indicate optimal models per criterion. L = linked-BL; UL = unlinked-BL.

Partitioned Models (L/UL)	-ln(*L*)	Parameters (n_{BL})	AIC	*w*	AIC_c	w_c	BIC	w_{BIC}	DT-Risk
M^1	43094.04	32 (23)	86252	0	86252	0	86471	0	1.04109
$M^2_{1+2,3}$ (L)	40141.04	41 (23)	80364	0	80365	0	80645	0	0.05792
$M^2_{1+3,2}$ (L)	41693.28	41 (23)	83469	0	83469	0	83749	0	0.73047
$M^3_{1,2,3}$ (L)	39835.26	50 (23)	79771	0	79771	0	**80113**	**1**	**0**
M^5_{gene} (L)	42947.97	68 (23)	86031	0	86033	0	86497	0	1.01059
$M^6_{1+2,3}$ (L)	41545.74	77 (23)	83245	0	83247	0	83773	0	0.67179
$M^6_{1+3,2}$ (L)	39942.33	77 (23)	80039	0	80040	0	80566	0	0.03720
$M^{10}_{1+2,3}$ (L)	39870.53	113 (23)	79967	0	79971	0	80740	0	0.06247
$M^{10}_{1+3,2}$ (L)	41445.56	113 (23)	83117	0	83121	0	83891	0	0.66555
$M^{11}_{1,2,3}$ (L)	39662.90	122 (23)	79570	0	79574	0	80405	0	0.06614
$M^{11}_{1,2,3}$ (L)	39598.59	122 (23)	79441	0	79446	0	80276	0	0.12056
$M^{11}_{1,2,3}$ (L)	39569.18	122 (23)	79382	0	79387	0	80217	0	0.08300
$M^{15}_{1,2,3}$ (L)	39497.13	158 (23)	**79310**	**1**	**79318**	**1**	80392	0	0.13652
$M^2_{1+2,3}$ (UL)	40067.02	64 (46)	80262	0	80263	0	80700	0	0.59139
$M^2_{1+3,2}$ (UL)	41644.11	64 (46)	83416	0	83417	0	83854	0	0.75580
$M^3_{1,2,3}$ (UL)	39737.41	96 (69)	79667	0	79670	0	80324	0	0.59523
M^5_{gene} (UL)	42866.18	142 (115)	86052	0	86060	0	87147	0	1.00447
$M^6_{1+2,3}$ (UL)	41419.74	192 (138)	83223	0	83234	0	84538	0	0.63647
$M^6_{1+3,2}$ (UL)	39782.87	192 (138)	79950	0	79961	0	81264	0	0.59516
$M^{10}_{1+2,3}$ (UL)	39655.76	320 (230)	79952	0	79983	0	82142	0	0.96482
$M^{10}_{1+3,2}$ (UL)	41275.54	320 (230)	83191	0	83222	0	85381	0	0.63611
$M^{11}_{1,2,3}$ (UL)	39466.00	352 (253)	79636	0	79674	0	82045	0	0.96659
$M^{11}_{1,2,3}$ (UL)	39366.16	352 (253)	79436	0	79474	0	81846	0	0.97320
$M^{11}_{1,2,3}$ (UL)	39349.36	352 (253)	79403	0	79440	0	81812	0	0.60394
$M^{15}_{1,2,3}$ (UL)	39222.19	480 (345)	79404	0	79476	0	82690	0	0.97470

Table 5.8 Identifying optimal partitioned-models for the paleognath alignment using marginal model likelihoods. BL priors are denoted as BL-X, where X is the exponential rate parameter. H.M. –ln(L) is the estimated harmonic mean model log likelihood. Pr(T(X)) is the posterior predictive p-value of the realized empirical multinomial test statistic. D_{LI} is the variance penalized distance between the realized empirical test statistic and the mean of the predictive distribution of the statistic. Values in bold represent the optimal partitioned-model for a given alignment/BL-prior determined from Bayes factors and D_{LI}. * indicates that the model is rejected at the $\alpha = 0.05$ level (two-sided). L = linked-BL; UL = unlinked-BL.

	Parameters	**BL-10**			**BL-20**			**BL-100**		
Model	**(n_{BL})**	**H.M. –ln(L)**	**Pr(T(X))**	**D_{LI}**	**H.M. –ln(L)**	**Pr(T(X))**	**D_{LI}**	**H.M. –ln(L)**	**Pr(T(X))**	**D_{LI}**
M^{1}	33 (23)	42796.72	0.944	808.22	42805.07	0.960	865.20	42865.64	0.998*	1336.49
$M^{2}_{1+2,3}$ (L)	43 (23)	40044.67	0.196	402.42	40057.22	0.237	377.55	40147.16	0.569	311.14
$M^{2}_{1+3,2}$ (L)	43 (23)	41373.45	0.920	654.46	41386.63	0.946	721.17	41459.45	0.998*	1162.90
$M^{3}_{1,2,3}$ (L)	53 (23)	39709.52	0.218	382.05	39713.08	0.251	364.01	39814.57	0.586	309.53
M^{5}_{gene} (L)	73 (23)	42665.90	0.930	766.29	42672.44	0.951	824.85	42727.11	0.999*	1337.01
$M^{6}_{1+2,3}$ (L)	83 (23)	41268.02	0.976*	830.96	41276.72	0.985*	892.44	41377.74	1.000*	1323.44
$M^{6}_{1+3,2}$ (L)	79 (23)	39858.96	0.188	402.94	39882.14	0.216	383.05	39960.75	0.545	303.60
$M^{10}_{1+2,3}$ (L)	122 (23)	39812.95	0.211	389.32	39826.66	0.257	363.61	39916.37	0.569	309.06
$M^{10}_{1+3,2}$ (L)	114 (23)	41231.40	0.665	415.44	41240.17	0.741	453.33	41317.54	0.975	838.62
$M^{11}_{1,2,3}$ (L)	122 (23)	39602.34	0.199	399.18	39613.58	0.241	373.39	39707.81	0.561	305.08
$M^{11}_{1,2,3}$ (L)	129 (23)	39507.85	0.236	**362.27**	39526.23	0.272	347.63	39619.84	0.608	309.82
$M^{11}_{1,2,3}$ (L)	120 (23)	39511.80	0.198	384.82	39508.44	0.204	384.29	39602.81	0.495	298.12
$M^{15}_{1,2,3}$ (L)	159 (23)	**39453.51**	0.171	421.67	**39472.57**	0.215	379.66	**39563.82**	0.561	**296.34**
$M^{2}_{1+2,3}$ (UL)	66 (46)	40031.64	0.145	442.72	40073.31	0.213	394.00	40286.13	0.845	448.07
$M^{2}_{1+3,2}$ (UL)	66 (46)	41382.42	0.876	580.02	41396.03	0.928	681.20	41502.51	1.000*	1344.55
$M^{3}_{1,2,3}$ (UL)	99 (69)	39697.24	0.091	508.93	39740.99	0.163	425.57	39958.20	0.939	561.12
M^{5}_{gene} (UL)	165 (115)	42651.87	0.933	766.50	42683.78	0.991*	1091.76	42919.92	1.000*	3251.71
$M^{6}_{1+2,3}$ (UL)	198 (138)	41293.73	0.977*	839.95	41354.33	0.998*	1190.01	41722.17	1.000*	3666.02
$M^{6}_{1+3,2}$ (UL)	194 (138)	39840.07	0.101	772.70	39875.88	0.059	557.50	40087.70	0.995*	865.01
$M^{10}_{1+2,3}$ (UL)	329 (230)	39908.63	0.035	658.71	40033.22	0.135	466.24	40648.88	1.000*	2815.89
$M^{10}_{1+3,2}$ (UL)	321 (230)	41312.97	0.277	448.37	41335.51	0.731	453.29	41694.19	1.000*	3250.18
$M^{11}_{1,2,3}$ (UL)	352 (253)	39797.54	0.005*	891.72	39865.70	0.063	581.08	40472.55	1.000*	2524.02
$M^{11}_{1,2,3}$ (UL)	359 (253)	39613.60	0.042	621.49	39734.86	0.292	**343.40**	40401.78	1.000*	3054.02
$M^{11}_{1,2,3}$ (UL)	350 (253)	39536.46	0.000*	1186.85	39572.17	0.004*	853.14	39799.24	1.000*	1154.32
$M^{15}_{1,2,3}$ (UL)	481 (345)	39596.92	0.000*	1119.59	39692.28	0.025	675.04	40328.22	1.000*	2961.03

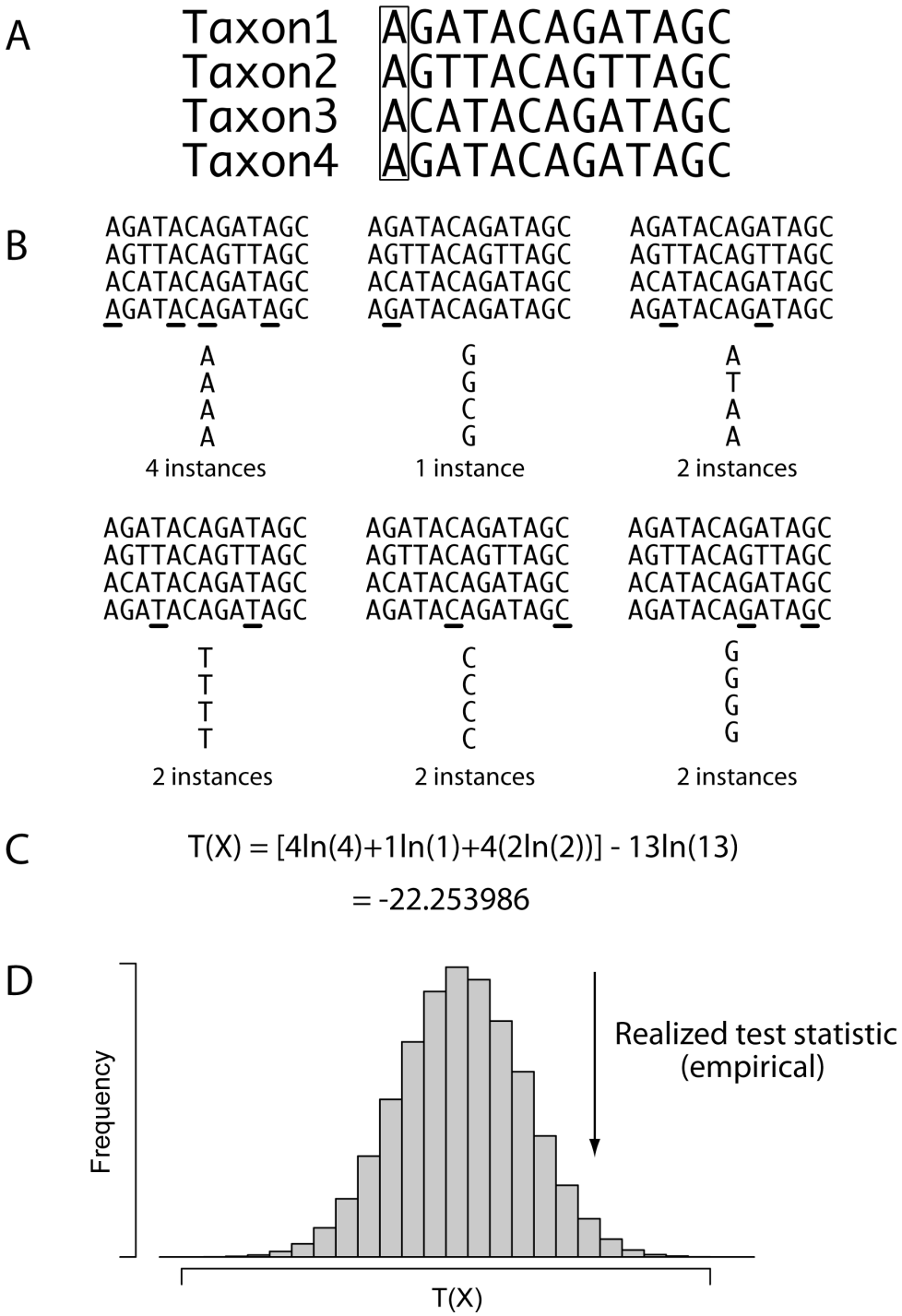

Figure 5.1 Calculation of the multinomial test statistic summarizing the shape of the alignment column frequency spectrum. (A) A posterior predictive nucleotide alignment (the same length of the empirical) is simulated from posterior parameter sample values. What is of interest are the alignment columns (the first of which is boxed), which represent taxon states for homologous characters. (B) Summary of the composition of the alignment column frequency spectrum begins with enumerating the distinct column patterns. (C) The multinomial test statistic summarizes the number and relative frequencies of distinct alignment columns. (D) Simulating many predictive alignments generates a distribution of the test statistic which is compared to the realized test statistic from the empirical alignment (arrow).

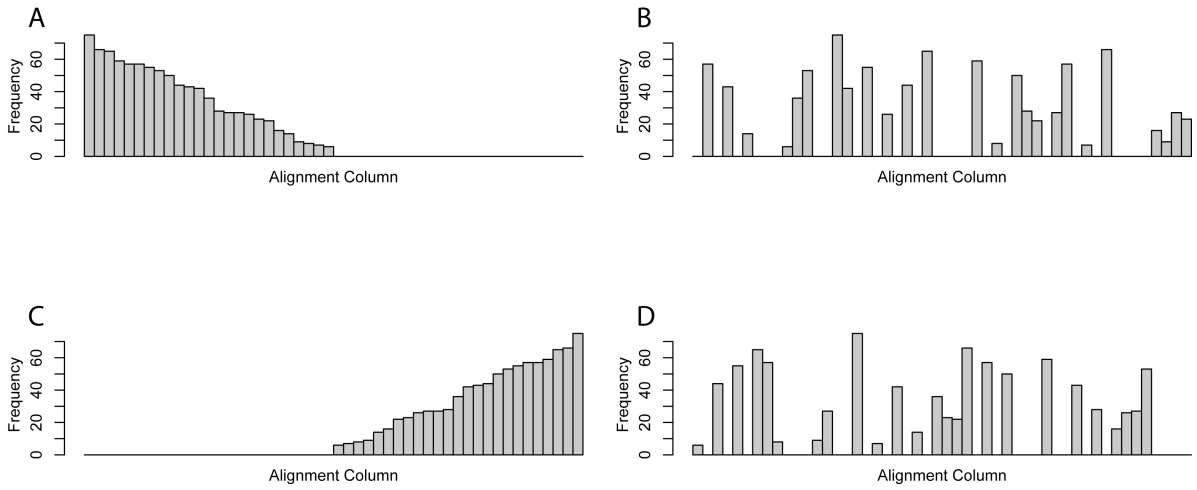

Figure 5.2 Hypothetical alignment column frequency spectra. Each column represents a distinct alignment column pattern, and the ordering of patterns across plots is consistent (i.e. patterns are labelled). Plots (A) and (C) have no patterns in common; similar with plots (B) and (D). Although plots (A) and (B) share roughly 50% of site patterns, the relative frequencies of the labelled patterns differs dramatically; similar with plots (C) and (D). Nevertheless, all four spectra generate the same multinomial test statistic ($T(X) = -2787.316839$) because the unlabelled spectra are identical (i.e. the number and relative frequencies of unlabelled column patterns is constant).

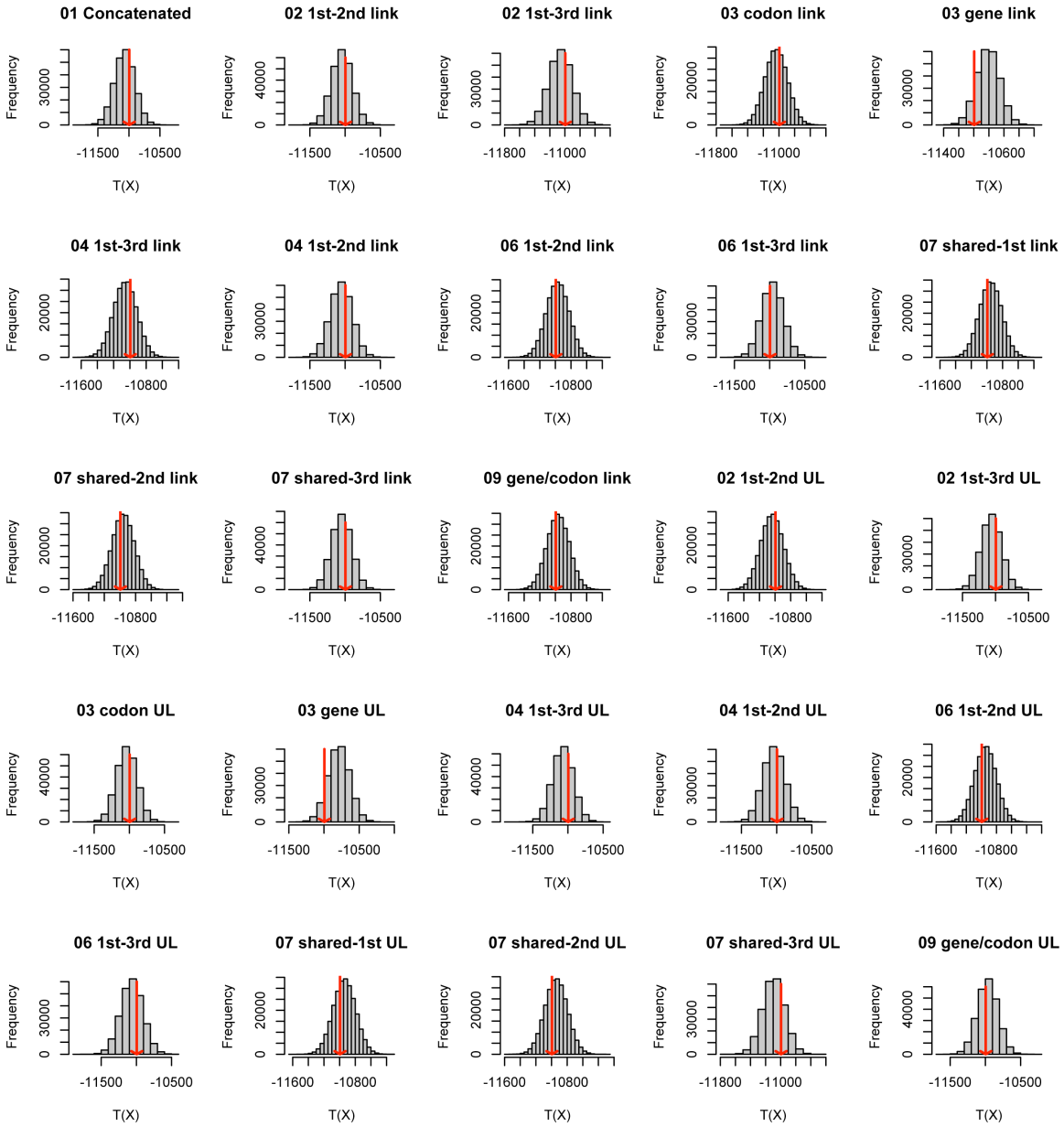

Figure 5.3 Model adequacy results for the primate alignment assuming an exp(10) prior on branch length parameters. Red arrows indicate the value of the realized test statistic from the empirical matrix. 'Link' indicates that branches were constrained to be proportional in the analysis model, while 'UL' indicates that the analyzing model made no assumptions regarding branch length proportionality across partitions.

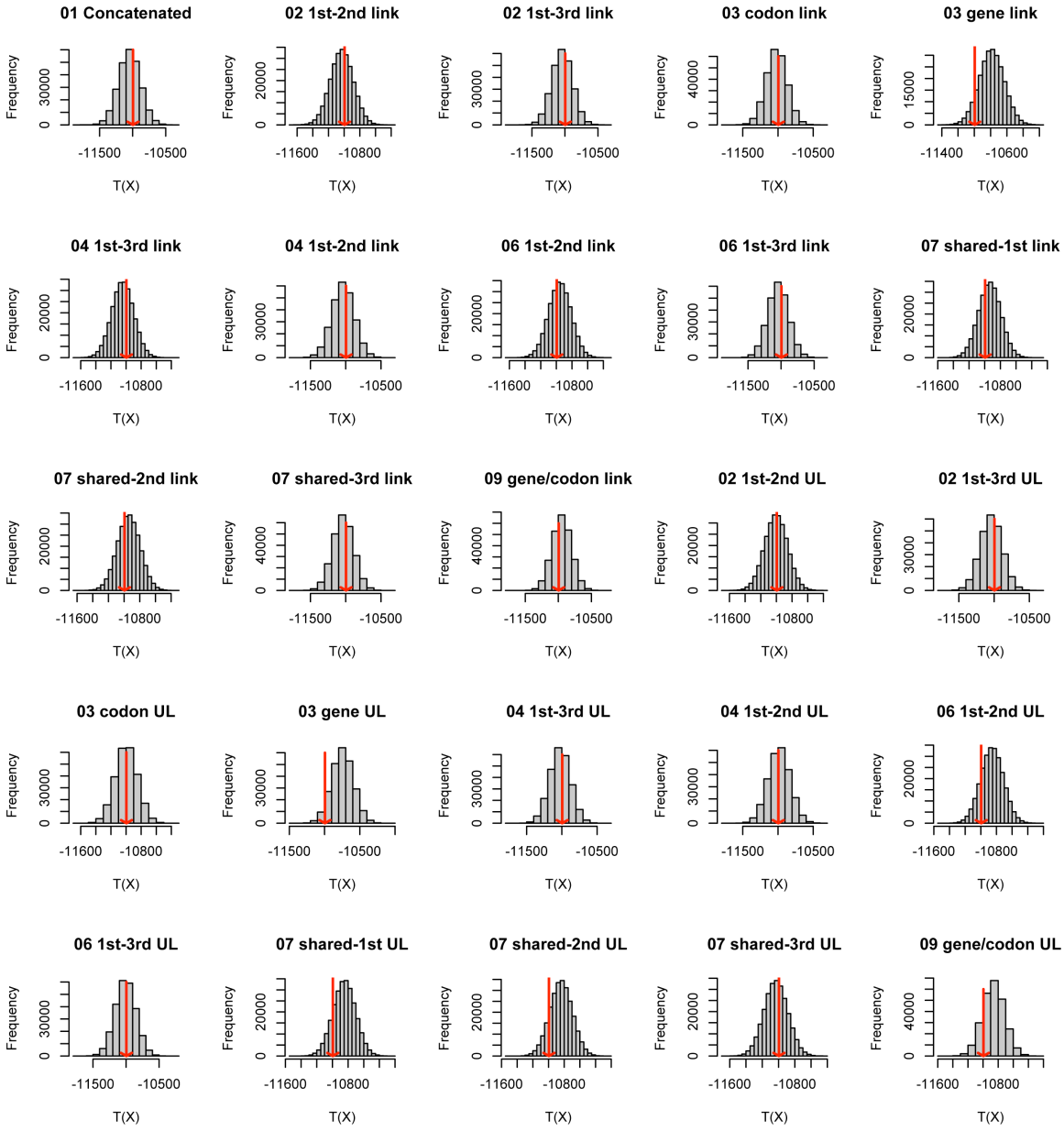

Figure 5.4 Model adequacy results for the primate alignment assuming an exp(20) prior on branch length parameters. Red arrows indicate the value of the realized test statistic from the empirical matrix. 'Link' indicates that branches were constrained to be proportional in the analysis model, while 'UL' indicates that the analyzing model made no assumptions regarding branch length proportionality across partitions.

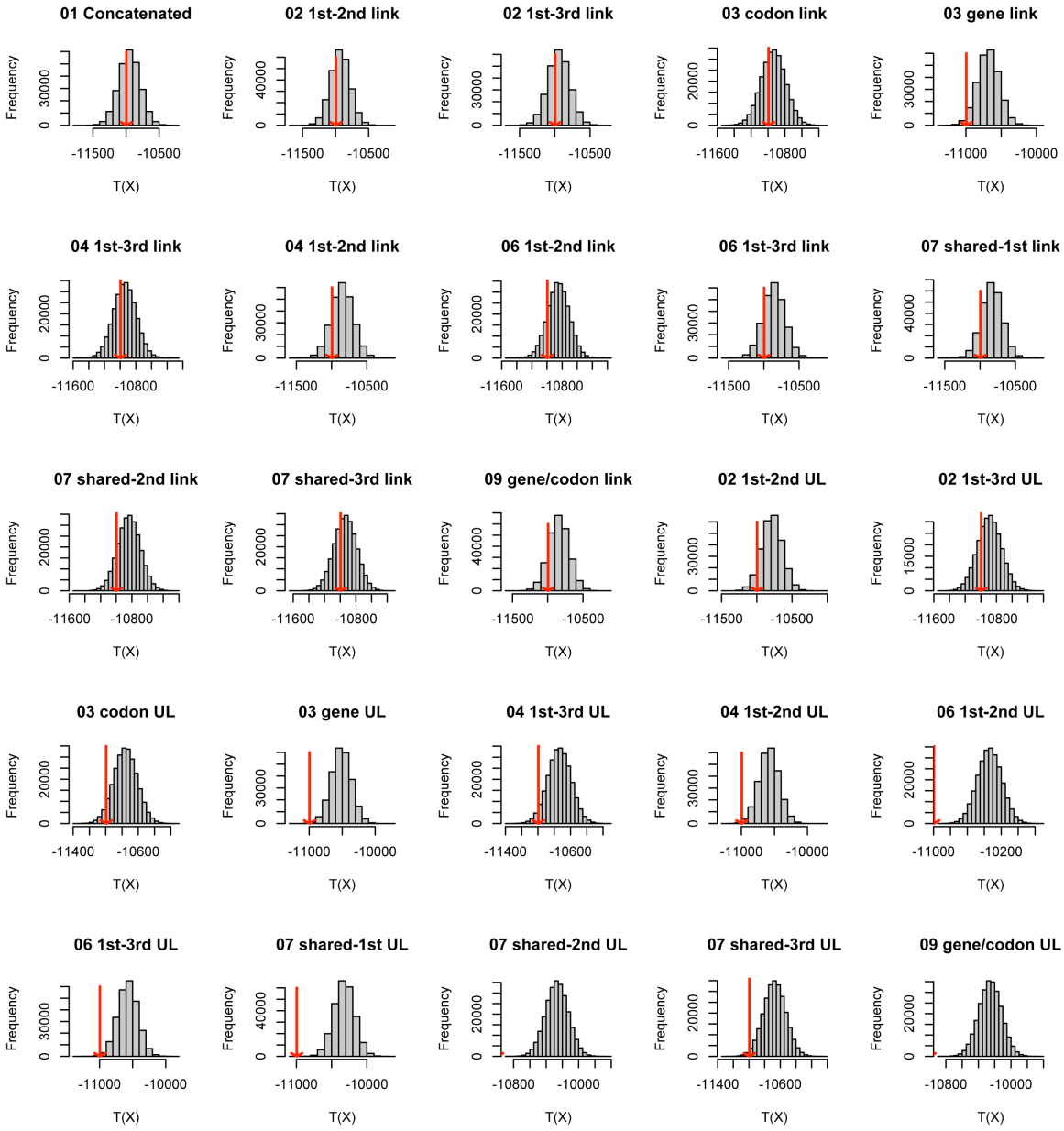

Figure 5.5 Model adequacy results for the primate alignment assuming an exp(100) prior on branch length parameters. Red arrows indicate the value of the realized test statistic from the empirical matrix. 'Link' indicates that branches were constrained to be proportional in the analysis model, while 'UL' indicates that the analyzing model made no assumptions regarding branch length proportionality across partitions.

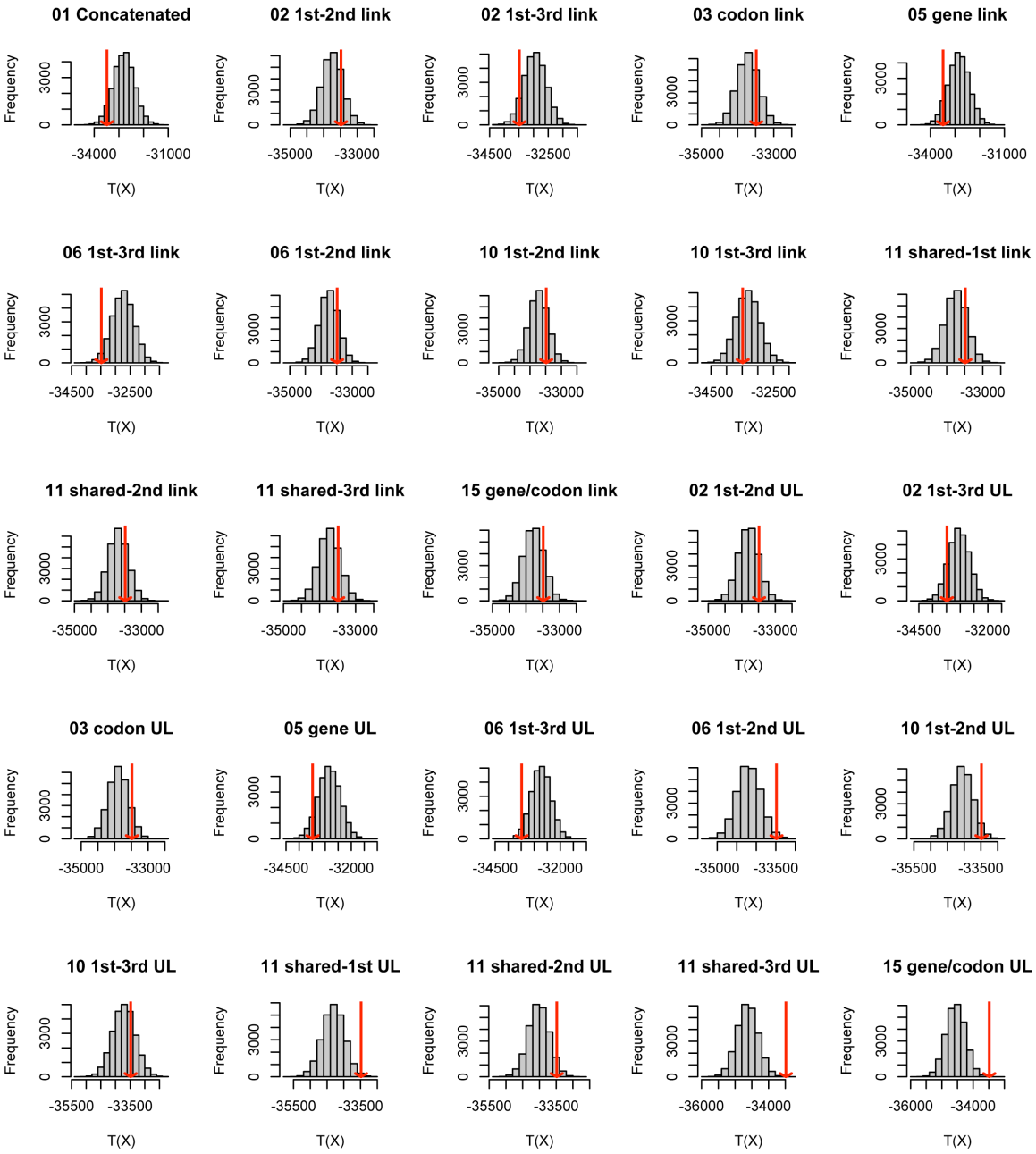

Figure 5.6 Model adequacy results for the paleognath alignment assuming an exp(10) prior on branch length parameters. Red arrows indicate the value of the realized test statistic from the empirical matrix. 'Link' indicates that branches were constrained to be proportional in the analysis model, while 'UL' indicates that the analyzing model made no assumptions regarding branch length proportionality across partitions.

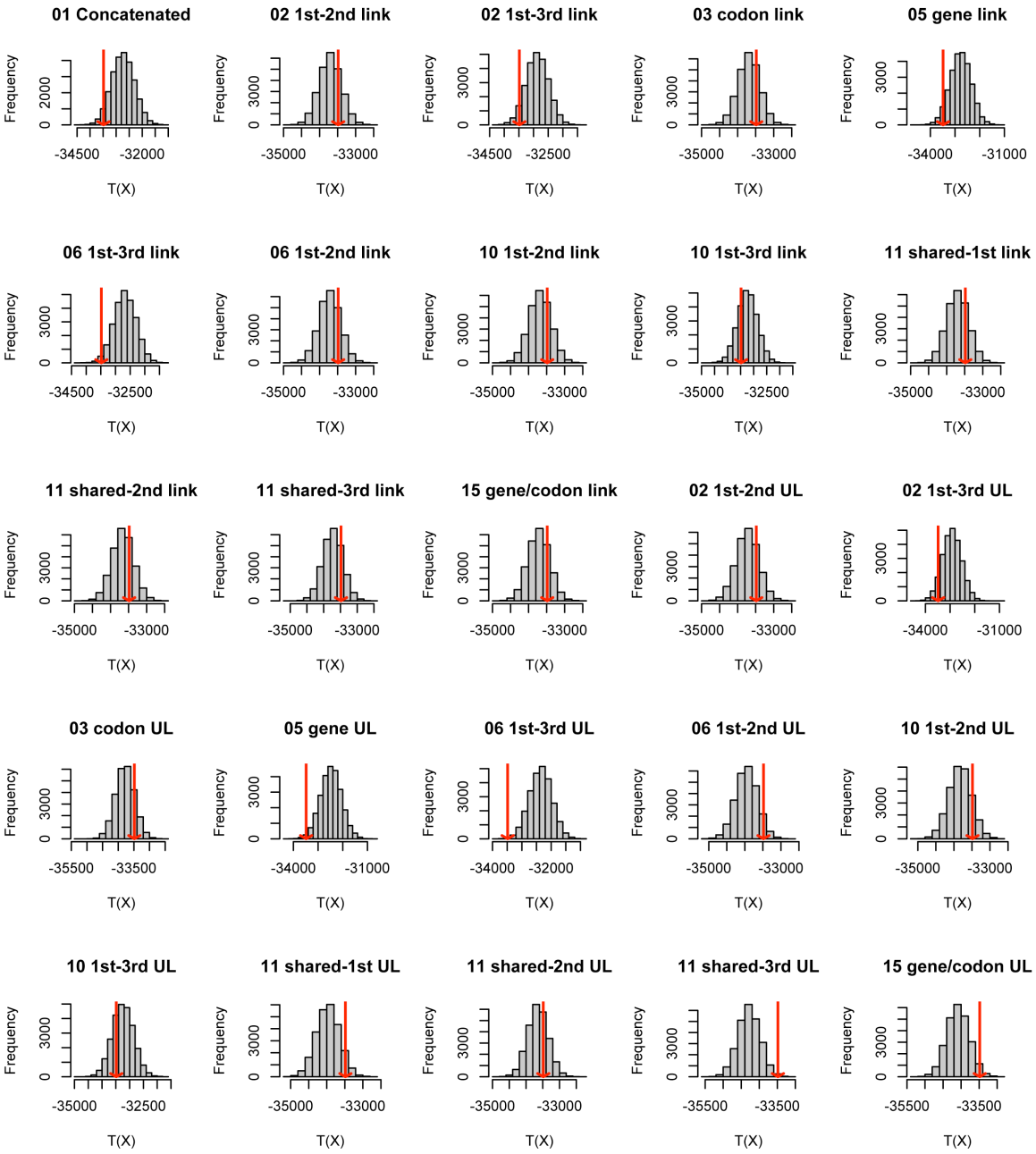

Figure 5.7 Model adequacy results for the paleognath alignment assuming an exp(20) prior on branch length parameters. Red arrows indicate the value of the realized test statistic from the empirical matrix. 'Link' indicates that branches were constrained to be proportional in the analysis model, while 'UL' indicates that the analyzing model made no assumptions regarding branch length proportionality across partitions.

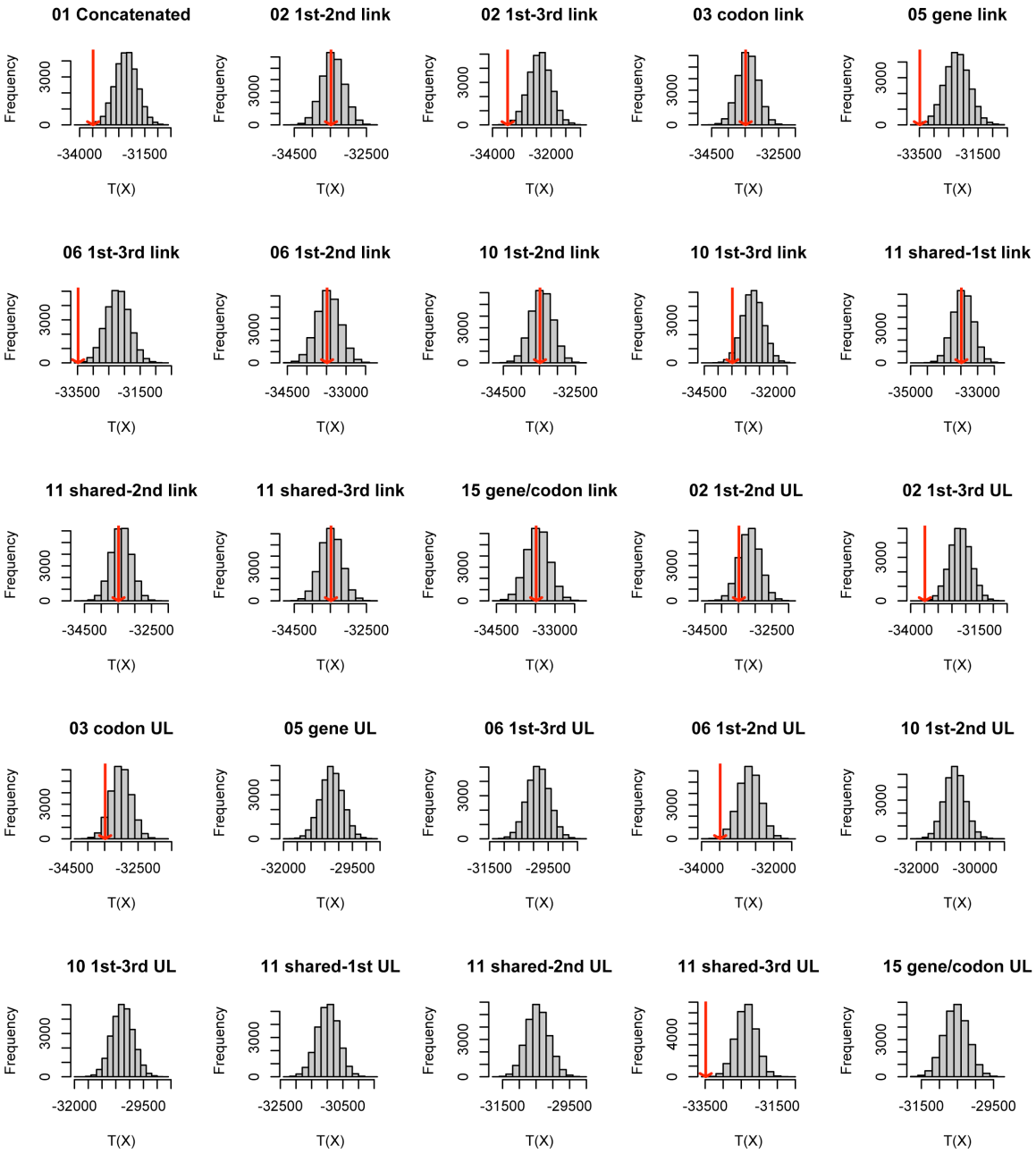

Figure 5.8 Model adequacy results for the paleognath alignment assuming an exp(100) prior on branch length parameters. Red arrows indicate the value of the realized test statistic from the empirical matrix. 'Link' indicates that branches were constrained to be proportional in the analysis model, while 'UL' indicates that the analyzing model made no assumptions regarding branch length proportionality across partitions.

REFERENCES

Abdo, Z., V. Minin, P. Joyce, and J. Sullivan. 2005. Accounting for uncertainty in the tree topology has little effect on the decision-theoretic approach to model selection in phylogeny estimation. Molecular Biology and Evolution 22:691-703.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267-281 *in* Second International Symposium on Information Theory (P. N. Petrov, and F. Csaki, eds.). Akademiai Kiado, Budapest.

Berlin, S., and H. Ellegren. 2001. Clonal inheritance of avian mitochondrial DNA. Nature 413:37-38.

Bollback, J. P. 2002. Bayesian model adequacy and choice in phylogenetics. Molecular Biology and Evolution 19:1171-1180.

Bos, D. H., and D. Posada. 2005. Using models of nucleotide evolution to build phylogenetic trees. Developmental & Comparative Immunology 29:211-227.

Brandley, M. C., A. Schmitz, and T. W. Reeder. 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Systematic Biology 54:373-390.

Brown, J. W., J. A. Johnson, and D. P. Mindell. to be submitted-a. Treatment of branch length parameters in partitioned phylogenetic models: an empirical case study with New World vultures (Aves: Cathartidae). Systematic Biology.

Brown, J. W., Y. Liu, Y.-L. Qiu, and Z. Yang. to be submitted-b. A molecular genetic timescale of early land plant diversification. Proceedings of the National Academy of Sciences of the United States of America.

Buckland, S. T., K. P. Burnham, and N. H. Augustin. 1997. Model selection: an integral part of inference. Biometrics 53:603-618.

Buckley, T. R. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Systematic Biology 51:509-523.

Buckley, T. R., and C. W. Cunningham. 2002. The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. Molecular Biology and Evolution 19:394-405.

Buckley, T. R., C. Simon, and G. K. Chambers. 2001. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. Systematic Biology 50:67-86.

Burnham, K. P., and D. J. Anderson. 2004. Mulitmodel inference: understanding AIC and BIC in model selection. Sociological Methods Research 33:261-304.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Second edition. Springer-Verlag, New York.

Caterino, M. S., R. D. Reed, M. M. Kuo, and F. A. H. Sperling. 2001. A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilonidae). Systematic Biology 51:106-127.

Cooper, A., C. Lalueza-Fox, S. Anderson, A. Rambaut, J. J. Austin, and R. Ward. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature 409:704-707.

Cummings, M. P., S. P. Otto, and J. Wakeley. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Molecular Biology and Evolution 12:814-822.

DeBry, R. W. 1999. Maximum likelihood analysis of gene-based and structure-based process partitions, using mammalian mitochondrial genomes. Systematic Biology 48:286-299.

Erixon, P., B. Svennblad, T. Britton, and B. Oxelman. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Systematic Biology 52:665-673.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368-376.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Inc., Sunderland, Massachusetts.

Gaut, B. S., and P. O. Lewis. 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. Molecular Biology and Evolution 12:152-161.

Gelfand, A. E., and S. K. Ghosh. 1998. Model choice: A minimum posterior predictive loss approach. Biometrika 85:1-11.

Goldman, N. 1993. Statistical tests of models of DNA substitution. Journal of Molecular Evolution 36:182-198.

Griffiths, C. S. 1997. Corrleation of functional domains and rates of nucleotide substitution in cytochrome b. Molecular Phylogenetics and Evolution 7:352-365.

Gu, X., Y.-X. Fu, and W.-H. Li. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Molecular Biology and Evolution 12:546-557.

Haddrath, O., and A. J. Baker. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. Proceedings of the Royal Society of London B Biological Sciences 268:939-945.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Harshman, J., E. L. Braun, M. J. Braun, C. J. Huddleston, R. C. K. Bowie, J. L. Chojnowski, S. J. Hackett, K.-L. Han, R. T. Kimball, B. D. Marks, K. J. Miglia, W. S. Moore, S. Reddy, F. H. Sheldon, D. W. Steadman, S. J. Steppan, C. C. Witt, and T. Yuri. 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. Proceedings of the National Academy of Sciences 105:13462-13467.

Hasegawa, M., H. Kishino, and T.-a. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22:160-174.

Holder, M., and P. O. Lewis. 2003. Phylogeny estimation: traditional and Bayesian approaches. Nature Reviews Genetics 4:275-284.

Holder, M. T., D. J. Zwickl, and C. Dessimoz. 2008. Evaluating the robustness of phylogenetic methods to among-site variability in substitution processes. Philosophical Transactions of the Royal Society of London B Biological Sciences 363:4013-4021.

Huelsenbeck, J. P., B. Larget, and M. E. Alfaro. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. Molecular Biology and Evolution 21:1123-1133.

Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. Trends in Ecology & Evolution 19:101-108.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association 90:773-795.

Kelsey, C. R., K. A. Crandall, and A. F. Voevodin. 1999. Different models, different trees: the geographic origin of PTLV-I. Molecular Phylogenetics and Evolution 13:336-347.

Kuha, J. 2004. AIC and BIC: comparisons of assumptions and performance. Sociological Methods & Research 33:188-229.

Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. Annals of Statistics 22:79-86.

Lanave, C., G. Preparata, C. Sacone, and G. Serio. 1984. A new method for calculating evolutionary substitution rates. Journal of Molecular Evolution 20:86-93.

Laud, P. W., and J. G. Ibrahim. 1995. Predictive model selection. Journal of the Royal Statistical Society. Series B (Methodological) 57:247-262.

Liò, P., and N. Goldman. 1998. Models of molecular evolution and phylogeny. Genome Research 8:1233-1244.

Maddison, W. P. 1997. Gene trees in species trees. Systematic Biology 46:523-536.

Minin, V., Z. Abdo, P. Joyce, and J. Sullivan. 2003. Performance-based selection of likelihood models for phylogenetic estimation. Systematic Biology 52:674-683.

Naylor, G. J. P., and W. M. Brown. 1997. Structural biology and phylogeny estimation. Nature 388:527-528.

Naylor, G. J. P., T. M. Collins, and W. M. Brown. 1995. Hydrophobicity and phylogeny. Nature 373:565-566.

Nylander, J. A. A. 2004. MrModeltest v2, version 2.3. Program distributed by author. Evolutionary Biology Centre, Uppsala University.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53:47-67.

Phillips, M. J., G. C. Gibb, E. A. Crimp, and D. Penny. 2010. Tinamous and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. Systematic Biology 59:90-107.

Pol, D. 2004. Empirical problems of the hierarchical likelihood ratio test for model selection. Systematic Biology 53:949-962.

Posada, D. 2008. jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25:1253-1256.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53:793-808.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817-818.

Posada, D., and K. A. Crandall. 2001a. Selecting the best-fit model of nucleotide substitution. Systematic Biology 50:580-601.

Posada, D., and K. A. Crandall. 2001b. Simple (wrong) models for complex trees: A case from Retroviridae. Molecular Biology and Evolution 18:271-275.

Pupko, T., D. Huchon, N. Okada, and M. Hasegawa. 2002. Combining multiple data sets in a likelihood analysis: Which models are the best? Molecular Biology and Evolution 19:2294-2307.

Rambaut, A., and A. J. Drummond. 2003. Tracer, version 1.3. Available from the authors (http://evolve.zoo.ox.ac.uk).

Rambaut, A., and N. C. Grassly. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Computer Applications in the Biosciences 13:235-238.

Rannala, B. 2002. Identifiability of parameters in MCMC Bayesian inference of phylogeny. Systematic Biology 51:754-760.

Reed, R. D., and F. A. H. Sperling. 1999. Interaction of process partitions in phylogenetic analysis: an example from the swallowtail butterfly genus Papilo. Molecular Biology and Evolution 16:286-297.

Rodríguez, F., J. L. Oliver, A. Marín, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. Journal of Theoretical Biology 142:485-501.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Russo, C. A., N. Takezaki, and M. Nei. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Molecular Biology and Evolution 13:525-536.

Saitou, N., and M. Nei. 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Sanderson, M. J., and J. Kim. 2000. Parametric phylogenetics? Systematic Biology 49:817-829.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6:461-464.

Shapiro, B., A. Rambaut, and A. J. Drummond. 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Molecular Biology and Evolution 23:7-9.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Steel, M. 2005. Should phylogenetic models be trying to 'fit an elephant'? Trends in Genetics 21:307-309.

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

Sullivan, J., and P. Joyce. 2005. Model selection in phylogenetics. Annual Review of Ecology, Evolution, and Systematics 36:445-466.

Sullivan, J., and D. L. Swofford. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. Journal of Mammalian Evolution 4:77-86.

Sullivan, J., and D. L. Swofford. 2001. Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? Systematic Biology 50:723-729.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4 (beta 10). Sinauer Associates, Sunderland, Massachusetts.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic Inference. Pages 407-514 *in* Molecular Systematics (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.

Swofford, D. L., and J. Sullivan. 2003. Phylogenetic inference using parsimony and maximum likelihood using PAUP*. Pages 160-196 *in* The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny (M. Salemi, and A. M. Vandamme, eds.). Cambridge University Press, Cambridge, UK.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences:57–86.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research 22:4673-4680.

Weakliem, D. L. 1999. A critique of the Bayesian information criterion for model selection. Sociological Methods & Research 27:359-397.

Wolfe, K. H., P. M. Sharp, and W.-H. Li. 1989. Mutation rates differ among regions of the mammalian genome. Nature 337:283-285.

Yang, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Molecular Biology and Evolution 10:1396-1401.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. Journal of Molecular Evolution 39:306-314.

Yang, Z. 1996a. Among-site rate variation and its impact on phylogenetic analyses. Trends in Ecology & Evolution 11:367-372.

Yang, Z. 1996b. Maximum-likelihood models for combined analyses of multiple sequence data. Journal of Molecular Evolution 42:587-596.

Zardoya, R., and A. Meyer. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Molecular Biology and Evolution 13:933-942.