Point of View

Syst. Biol. 59(1):108–117, 2010

DOI:10.1093/sysbio/syp080
Advance Access publication on November 17, 2009

Cryptic Failure of Partitioned Bayesian Phylogenetic Analyses: Lost in the Land of Long Trees

DAVID C. MARSHALL*

*Department of Ecology and Evolutionary Biology, University of Connecticut, 75 N. Eagleville Road, U-3043, Storrs, CT 06269, USA; *Correspondence to be sent to: Department of Ecology and Evolutionary Biology, University of Connecticut, 75 N. Eagleville Road, U-3043, Storrs, CT 06269, USA; E-mail: david.marshall@uconn.edu.*

Received 23 October 2008; reviews returned 25 November 2008; accepted 15 October 2009
Associate Editor: Frank E. Anderson

Partitioned Bayesian phylogenetic analyses of routine genetic data sets, constructed using MrBayes (Ronquist and Huelsenbeck 2003), can become trapped in regions of parameter space characterized by unrealistically long trees and distorted partition rate multipliers. Such analyses commonly fail to reach stationarity during hundreds of millions of generations of sampling—many times longer than most published analyses. Some data sets are so prone to this problem that paired MrBayes runs begun from different starting trees repeatedly find the same incorrect long-tree solutions and consequently pass the most commonly employed tests of stationarity, including the average standard deviation of split frequencies (ASDSF) and the potential scale reduction factor (PSRF) statistics offered by MrBayes (Gelman and Rubin 1992). In these situations, failure to reach stationarity is recognizable only in light of prior knowledge of model parameters, such as the expectation that third-codon-position sites usually evolve fastest in protein-coding genes. The conditions that lead to the long-tree problem are frequently encountered in phylogenetic studies today, and I present 6 demonstration examples from the literature. Although the effects on tree length (TL) are often dramatic, effects on topology appear to be subtle. Susceptibility to the problem is sometimes predicted by the difference between the true TL and the starting TL. In some cases, the problems described here can be avoided or reduced by manipulation of the starting TL and/or by adjustments to the prior on branch lengths. In more difficult situations, accurate branch length estimation may not be possible with Bayesian methods because of dependence of the solution on the branch length prior.

BAYESIAN MARKOV CHAIN MONTE CARLO ANALYSES AND STATIONARITY

Bayesian phylogenetic analysis, like other methods that employ Markov Chain Monte Carlo (MCMC) sampling, offers accurate estimation of parameter values and posterior probabilities of topologies only if the chains are run for a sufficiently long time at "stationarity" (Tierney 1994; Huelsenbeck and Ronquist 2005; Ronquist et al. 2005). (Note that 1 Bayesian MCMC "run" typically includes multiple Metropolis-coupled Markov chains, several "heated" and one "cold," hence the term MCMCMC or MC^3.) In practical terms, stationarity is reached some number of generations after the initiation of an MC^3 run, at the point when the starting parameter values no longer predict the current states of the chains. The prestationarity phase is known as the burn-in, and samples from this phase are discarded. A key challenge for all Bayesian applications is determining when a run has reached stationarity—the point when the investigator can begin to accurately estimate the posterior probability distribution.

Two general criteria are used for determining whether stationarity has been reached—stability of the solution over time and reproducibility of the solution from different starting conditions. Plots of the model log-likelihood and of the separate parameters, over time, are examined visually for a given run to confirm stable mean values over millions of generations (sometimes tens of millions). In addition, multiple runs are conducted, commonly from different starting trees, to confirm that the same solution is reached in each replicate. MrBayes v3.1, in its default mode, compares the parameter estimates and bipartition frequencies from paired runs (nruns = 2) using ASDSF and PSRF after exclusion of samples from the estimated burn-in period (see the MrBayes v3.1 manual as well as Gelman and Rubin 1992). The ASDSF value reflects the similarity of the topologies sampled by the 2 independent analyses, whereas PSRF compares their final mean parameter estimates (including posterior probabilities of nodes). PSRF values are accepted as they approach 1.000, and ASDSF values are often accepted once they drop below 0.01 (although this value is only an informal recommendation, and some studies adjust the critical level based on the number of taxa). Adequacy of the posterior sample size, an important third consideration, is commonly assessed through autocorrelation statistics

(e.g., as employed in Tracer: Rambaut and Drummond 2003).

Most studies report that stationarity is reached within a few hundred thousand generations of sampling. However, there have been indications that highly parameterized models and/or large data sets may require much longer analyses (e.g., Goloboff and Pol 2005; Miya et al. 2005; Soltis et al. 2007). These latter examples, together with the high complexity of modern substitution models and, especially, data partitioning schemes (Nylander et al. 2004; Brandley et al. 2005), raise the possibility that chains could become trapped in local optima for extended periods, leading to incorrect diagnosis of stationarity and incorrect parameter estimates. Indeed, signs of this problem have already been observed in analyses of both empirical data (Marshall et al. 2006, p. 998) and simulated data (Brown and Lemmon 2007, p. 646; Brown et al. 2010).

Branch Length Problems with Partitioned Bayesian Analyses—Empirical Data

I have found that replicate partitioned MrBayes analyses begun from different random starting trees can yield strikingly different estimates of the total amount of evolution, or TL (measured in substitutions/site), along with strongly contrasting estimates of relative partition rates (m_i, for the ith data partition). Figure 1 and Table 1 show the results of 2 such analyses of a 51-taxon, 2152 bp mitochondrial protein-coding data set from cicadas (genus *Kikihia*: Marshall et al. 2008; TreeBase accession number SN4716), begun from the same nexus file. Both analyses were run for approximately 680 million generations—much longer than most published studies. To achieve a very large posterior sample as fast as possible (in this case, 82 d), just 1 MC^3 run with 4 chains was completed in each analysis. The data were partitioned by codon position, and a separate general time-reversible + Γ model (Yang 1994a, 1994b) was used for each partition (see supplementary Appendix 1, available from http://www.sysbio.oxfordjournals.org). Substitution and among-site rate variation models for each partition were chosen using Modeltest v3.7 and the AIC criterion (Posada and Crandall 1998), except that $p_{\text{invar}}$was removed from the final models to reduce parameter correlation (Sullivan et al. 1999). The appropriateness of the 3-partition model was checked using the criterion 2 ln Bayes factor (2 ln BF) >10 (Kass and Raftery 1995; Nylander et al. 2004; Brandley et al. 2005) (2 ln BF was 143.16 in favor of the 3-partition model over a 2-partition alternative; see supplementary Appendix 1). Uncorrected (p) genetic divergence values for this data set (ranging up to 11.5%) are consistent with a mainly Plio-Pleistocene insect radiation. Analysis I found the superior log-likelihood score and a TL consistent with a 51-taxon data set of this depth (TL = 1.50), whereas Analysis II estimated a value over 5 times greater, TL = 7.70 (note scale bars in Fig. 1b). Estimates

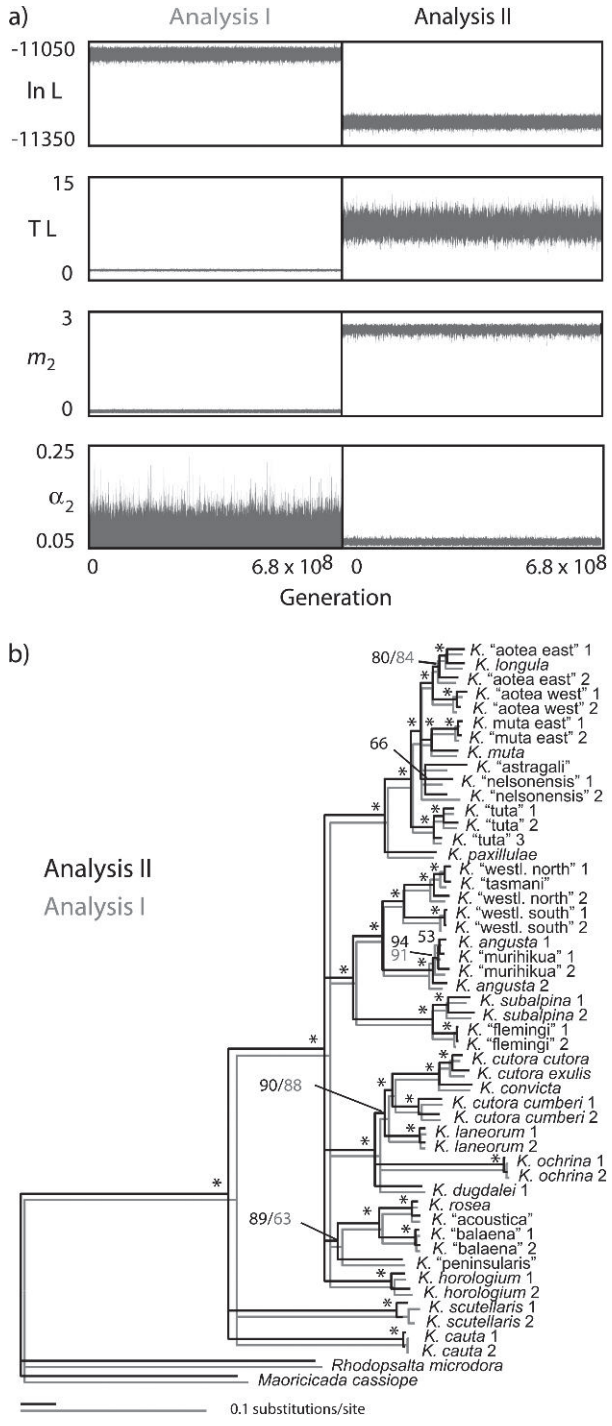

FIGURE 1. a) Model log-likelihoods (ln L) and selected parameter estimates found during replicated 680 million generation MrBayes analyses that reached contrasting solutions. Analysis I estimated TL to be 1.50, whereas Analysis II became trapped on a long-tree solution, TL = 7.70, for the entire run. α_2 and m_2 are the second-codon-position among-site rate variation and partition rate multiplier parameters, respectively. The first 1% of the samples have been discarded as "burn-in." b) Overlapped consensus phylograms from the 2 analyses in Figure 1a. The scale bars indicate the difference in TL between the 2 results. Asterisks indicate Bayesian posterior probabilities of 98% or higher for both trees.

TABLE 1. Contrasting parameter estimates of replicate 680 million generation MrBayes analyses (Analyses I and II, Fig. 1), showing different TL solutions

Analysis	Ln L	TL	m_1	m_2	m_3	α_1	α_2	α_3	R_{CT}	R_{CG}	R_{AT}	R_{AG}
I	−11085.13	1.503	0.63	0.13	2.25	0.109	0.071	1.42	41.4	3.32	0.67	28.2
	−11,100	1.37	0.62	0.13	2.25	0.108	0.066	1.40	25.2	1.30	0.38	17.7
	−11,070	1.65	0.72	0.17	2.34	0.14	0.12	1.80	67.7	6.82	1.14	45.3
II	−11280.79	7.704	0.11	2.50	0.39	0.105	0.065	1.51	38.6	3.03	0.66	26.4
	−11,300	5.78	0.08	2.33	0.29	0.081	0.06	1.19	23.7	1.21	0.38	16.6
	−11,260	9.92	0.15	2.62	0.52	0.13	0.07	1.94	62.4	6.22	1.11	42.1

	R_{AC}	π_{A1}	π_{C1}	π_{G1}	π_{T1}	π_{A2}	π_{C2}	π_{G2}	π_{T2}	π_{A3}	π_{C3}	π_{G3}	π_{T3}
I	3.82	0.39	0.11	0.13	0.37	0.23	0.17	0.14	0.46	0.46	0.05	0.05	0.42
	2.19	0.35	0.09	0.11	0.34	0.20	0.14	0.12	0.43	0.43	0.05	0.05	0.41
	6.48	0.42	0.13	0.15	0.40	0.26	0.19	0.17	0.50	0.49	0.06	0.06	0.46
II	3.48	0.39	0.11	0.13	0.37	0.24	0.18	0.14	0.45	0.47	0.05	0.05	0.43
	2.01	0.35	0.09	0.11	0.34	0.21	0.16	0.11	0.41	0.43	0.04	0.04	0.40
	5.82	0.42	0.13	0.15	0.41	0.27	0.21	0.16	0.48	0.49	0.06	0.06	0.46

Notes: Ln L = harmonic mean log likelihood from MrBayes output; m_i = rate multipliers for each of the 3 mtDNA codon-position partitions; π_i = partition-specific base frequencies; α_i = partition-specific alpha values (gamma distribution hyperprior). R values are instantaneous nucleotide mutation rates (not partitioned). Means are given above and 95% confidence interval limits below. The first 1% of the samples in each analysis were discarded as burn-in.

of the relative partition rates also differed dramatically, with the "short-tree" analysis estimating a faster rate for the third-codon-position sites as expected, and the "long-tree" analysis finding a much faster rate for the second-position sites—an estimate at odds with empirical data on the evolution of protein-coding loci (Lin and Danforth 2004). Both analyses appeared stable in likelihood plots (Fig. 1a) and the consensus phylograms and branch support values were similar for the 2 runs (Fig. 1b). Remarkably, neither analysis sampled the area of parameter space inhabited by the other during more than half a billion generations of MC^3 sampling—an investigator examining either analysis alone would likely conclude that stationarity had been reached.

The problem of diagnosing stationarity in this situation becomes especially acute when a typical MrBayes v3.1 analysis is conducted with paired runs and both become stuck on the same long-tree solution, which happens easily with some data sets. In these cases not even the sensitive convergence diagnostics offered by MrBayes offer a clear indication that anything is amiss. Table 2 summarizes log-likelihood and key parameter estimates from 4 MrBayes v3.1 analyses of the *Kikihia* data set (nruns = 2 in each case). In 2 of the 4 analyses (Analyses IV and V), the paired replicates landed on the same solution involving extremely long trees and apparently nonsensical partition rate multipliers, but the convergence diagnostic statistics all fell well within accepted limits. Only 1 of the 8 independent MC^3 runs shown in Table 2 found the correct solution (Analysis VI, Run 1). Later in this paper, I will show that this problem is common in the phylogenetic literature.

BRANCH LENGTH PROBLEMS WITH PARTITIONED BAYESIAN ANALYSES—SIMULATED DATA

I have also demonstrated the TL problem under known simplified conditions using simulated data. Using Seq-Gen version 1.3.2 (Rambaut and Grassly 1997), and a randomly generated 25-taxon tree with simulated TL = 1.0 (using Phyl-O-Gen v1.1; Rambaut 2002), I generated one 2400 bp data set using the JC (Jukes and Cantor 1969) + Γ model for each of 3 equal-sized data partitions. I simulated between- and within-partition rate variation similar to that observed in the cicada mitochondrial DNA (mtDNA) data set by scaling the simulation TL to create relative partition rates of 0.7, 0.2, and 2.1, respectively, and by setting the simulated partition-specific α parameters to 0.4, 0.25, and 3.0, respectively. I used Modeltest version 3.7 (Posada and Crandall 1998) and its default settings to confirm the appropriateness of the simulated model for analysis of each partition under the AIC criterion, and I used the 2 ln BF >10 criterion to confirm the appropriateness of an analysis model partitioned by codon position (2 ln BF = 141.52 in favor of the 3-partition model over a 2-partition alternative; see supplementary Appendix 1). I then conducted 15 replicate MrBayes v3.1 analyses of this data set (with nruns = 1 in this case for speed) from random starting trees using the 3-way partitioned JC+Γ model with the among-site rate variation and base frequency parameters estimated separately for each partition (see supplementary Appendix 1).

As with the mitochondrial data set, the replicate analyses found strikingly different outcomes. During the first million generations, 8 analyses quickly converged on the correct solution—third partition fastest, second partition slowest, and a TL close to the simulated value of 1.0 (ln $L \approx -14608$). The other 7 initially converged on an incorrect long-tree solution—second partition fastest, third partition slowest, and TL near 4.0 (ln $L \approx -14837$). These analyses remained on the incorrect solution for 2.1, 2.6, 2.8, 11.1, 15.5, 16.2, and 20.6 million generations before each abruptly found the correct solution. (Bayesian analyses in many studies are terminated after 10 million generations or less.) In Table 3, results from 4 of these runs are summarized and

TABLE 2. Parameter estimates and convergence diagnostics illustrating the difficulty of diagnosing TL problems in partitioned Bayesian analyses of empirical data

	Analysis III ASDSF = 0.038373 Bipartition PSRF = 1.169			Analysis IV ASDSF = 0.001726 Bipartition PSRF = 1.001			Analysis V ASDSF = 0.001363 Bipartition PSRF = 1.000			Analysis VI ASDSF = 0.040448 Bipartition PSRF = 6.001		
	Run 1	Run 2	PSRF	Run 1	Run 2	PSRF	Run	Run 2	PSRF	Run 1	Run 2	PSRF
Ln L	−11194.5	−11256.8		−11255.9	−11269.5		−11260.5	−11259.4		−11061.4	−11254.7	
TL	7.629	8.480	1.274	8.432	8.474	1.001	8.453	8.447	1.000	1.524	8.514	9.176
m_1	0.106	0.095	1.199	0.096	0.095	1.001	0.095	0.095	1.000	0.582	0.095	12.576
m_2	2.498	2.550	1.248	2.547	2.550	1.001	2.549	2.548	1.000	0.115	2.552	53.020
m_3	0.395	0.353	1.241	0.355	0.353	1.001	0.354	0.354	1.000	2.306	0.352	36.375

Notes: Ln L = log likelihood; m_i = rate multipliers for each of the 3 mtDNA codon-position partitions; In each analysis, 1 or both of the paired MrBayes runs became trapped on incorrect long-tree solutions, as indicated by high TL estimates paired with multipliers indicating unexpected rapid evolution at second-codon-position sites (compare with the estimates for Analysis VI, Run 1, which found the correct solution). Convergence diagnostics offer no sign of the problem when both runs find the same incorrect solution (Analyses IV and V). All analyses used the same model and data set but were initiated with different starting trees.

results from runs with the same solution are compared using the PSRF statistic. Notably, the PSRF statistics observed while comparing any 2 short-tree and long-tree analyses were very similar, and all were very close to 1.000.

Interestingly, after each analysis eventually found the correct solution, subsequent samples were commonly taken from just 1 or 2 of the 4 independent Markov chains. This same pattern was observed in the analyses of empirical data. For example, in Analysis I above (Fig. 1, Table 1), which found the correct TL, all but one of the samples (Sample 2, generation 10,000) was taken from the same chain during the entire 680 million generation run—the remaining 3 chains were "left behind" in long-tree parameter space. In Analysis II, however, data were at least occasionally sampled from each of the 4 chains throughout the long analysis—meaning that all were sampling the same long-tree region of parameter space.

All of the simulated data analyses showed a correlation of long-tree solutions (high TLs) with unexpected values for the partition rate multipliers (m) and unusual α values for certain partitions. Analyses that yield short TL estimates always estimate a high m value for the third-codon-position partition and α values close to the simulated values. In contrast, failed analyses with high TL estimates usually estimate the second partition rate to be highest and a second partition α value that is extremely low. A very small value for α means a very high degree of among-site rate variation within a partition. Therefore, in the failed analyses, the comparatively limited number of variable sites in the second-position partition are modeled as evolving extremely quickly, increasing the average evolutionary rate of the partition dramatically despite the large number of invariant sites. Significantly, the partition-specific TLs for the data-rich third partition (obtained in each case by multiplying m_3 by the overall TL) are extremely similar in short-tree and long-tree analyses. The difference between the 2 solutions is mainly in the amount of evolution modeled for the data partition with the fewest variable sites.

CAUSES OF LONG-TREE SOLUTIONS

Why should an MC^3 analysis converge on such unrealistic m and α values for the second-codon-position data partition when the likelihood scores of these solutions are so much poorer than third-partition-fastest solutions? Examination of the early progress of these failed runs suggests a partial cause: MrBayes always begins its analyses with what is, for many data sets, a long initial starting tree, and this high TL value drives shifts in other model parameters that help commit the chain to a long-tree solution. TL begins at Generation 1 with a value equal to the number of branches in the (fully resolved) starting tree multiplied by the MrBayes initial branch length of 0.1. The number of branches in the fully resolved tree is equal to $(2 \times N) - 3$, where N is the number of taxa, so the initial TL equals 9.9 for the empirical data set in Analyses I-VI. As the chain

TABLE 3. Contrasting outcomes of replicate single-run MrBayes analyses of a 3-partition simulated data set

Parameter	Simulated value	Short-tree solutions			Long-tree solutions		
		Analysis VII mean	Analysis VIII mean	PSRF	Analysis IX mean	Analysis X mean	PSRF
Ln L	N/A	−14608.5	−14608.4	N/A	−14837.6	−14837.4	N/A
TL	1.0	1.011	1.011	1.000	4.342	4.360	1.001
m_1	0.7	0.725	0.726	1.000	0.169	0.167	1.002
m_2	0.2	0.191	0.190	1.000	2.342	2.349	1.003
m_3	2.1	2.083	2.084	1.000	0.489	0.484	1.003
α_1	0.4	0.512	0.513	1.000	0.506	0.508	1.000
α_2	0.25	0.353	0.330	1.000	0.085	0.085	1.000
α_3	3.0	4.377	4.449	1.002	4.448	4.474	1.004
A_1	0.25	0.232	0.232	1.000	0.233	0.233	1.000
C_1	0.25	0.264	0.264	1.000	0.264	0.263	1.000
G_1	0.25	0.240	0.240	1.000	0.240	0.240	1.001
T_1	0.25	0.264	0.264	1.000	0.264	0.264	1.000
A_2	0.25	0.257	0.257	1.000	0.256	0.256	1.000
C_2	0.25	0.247	0.247	1.001	0.246	0.247	1.000
G_2	0.25	0.246	0.246	1.000	0.247	0.245	1.000
T_2	0.25	0.250	0.251	1.000	0.251	0.251	1.000
A_3	0.25	0.246	0.245	1.001	0.245	0.246	1.000
C_3	0.25	0.249	0.249	1.000	0.249	0.249	1.000
G_3	0.25	0.247	0.248	1.001	0.248	0.247	1.000
T_3	0.25	0.258	0.257	1.000	0.258	0.258	1.000

Notes: Simulated values for each parameter are given on the left, followed by mean log-likelihoods and parameter estimates from 2 short-tree and 2 long-tree analyses. PSRF values for each parameter obtained by comparing the 2 short-tree or long-tree solutions are also given. Ln L = log likelihood; m_i = rate multipliers for each of the 3 mtDNA codon-position partitions; α_i = partition-specific alpha values (gamma distribution hyperprior). Other values are base frequencies. For Analyses IX and X, only the results from the first 20 and 16 million generations, respectively, are shown—the period before 1 or more MCMC chains abruptly located the correct solution (see text). The first 10% of the samples in each analysis were discarded as burn-in. Each analysis was begun from a different random starting tree.

progresses and topological improvements are made, TL becomes smaller as expected, but the initial value is so large that TL remains high for many generations. During this phase, substantial improvement to the overall likelihood score can be made by decreasing the rate multiplier for the data-rich third partition, which brings the partition-specific TL closer to its simulated value for that partition. However, because the rate multipliers must average 1, this necessitates an increase in the rate multiplier for at least one of the other slower partitions. Apparently the coordinated sharp decrease in α for the second partition ameliorates the worsening fit of the data in the second partition that is caused by the increasing rate multiplier. (Note that, in separate trials not shown here, increasing the number of gamma categories to 8 did not alleviate the problem.) Importantly, these "correlated" parameter shifts are all in a direction away from the true (simulated) values of those parameters, and they place the MCMC chain in what seems to be a local optimum in parameter space that can be difficult to escape despite its overall lower likelihood scores—the "land of long trees."

AVOIDING LONG-TREE LAND: STARTING BRANCH LENGTHS

The above hypothesis suggests that reducing the starting TL value might allow the analysis to find the short-tree solution before partition-specific α and m values can become too distorted. Therefore, using the simulated data set from Analyses VI–X above, I conducted 10 MrBayes v3.1 analyses under each of 5 different starting TL treatments, with each treatment using random starting trees containing branch lengths set for a particular starting TL value. MrBayes v3.1 offers no programmed method for changing starting parameter values—see supplementary Appendix 1 for details on how the starting tree and its length were designated. (Note that the starting parameter values are not the same as the Bayesian priors on those parameters.) To simplify the test and speed the trials, I monitored only 1 chain for each run and I ran each analysis for just 1 million generations, a sufficient period for examining the influence of starting TL.

The results confirmed the expected importance of starting TL value on the behavior of the MCMC chains. As the starting TL was adjusted closer to the simulated TL (Table 4), the chains became increasingly likely to find the correct short-tree solution during the trial period. With starting trees less than 2.33 in TL, all chains found the correct solution in less than 100,000 generations. Even extremely short starting branch lengths of 0.001 (TL = 0.047) yielded good results in which the chains quickly approached the simulated value for TL.

TABLE 4. Effect of starting TL on replicated MrBayes analyses of a simulated data set (1 chain per analysis)

Starting branch length	Starting TL	Number of correct solutions	Number of long-tree solutions
1.00	47	0	10
0.10	4.7	1	9
0.05	2.33	6	4
0.01	0.47	10	0
0.001	0.047	10	0

Notes: The MrBayes default starting branch length is 0.10. The "starting TL" is the starting mean branch length value multiplied by the number of branches in the fully resolved tree.

TABLE 5. Partition rate multipliers used in the 4 APRV treatments

	TL	Partition rate multipliers		
		1P	2P	3P
High APRV	1.0	2.1	0.2	0.7
Low APRV	1.0	0.85	0.5	1.55
No APRV	1.0	1.0	1.0	1.0
No APRV control	0.2	1.0	1.0	1.0

PARAMETER NONIDENTIFIABILITY AND THE LONG-TREE PROBLEM

In anecdotal trials, I have found that data sets vary in their susceptibility to the long-tree problem described above, perhaps because the local long-tree optimum does not exist or is not as sharply peaked for some combinations of data and model. Detailed exploration of the sources of the problem is beyond the scope of this paper (see Brown et al. 2010). However, I can briefly describe the results of 2 tests that suggest that the problem can be partly traced to the use of models that contain borderline nonidentifiable parameters (see Sullivan and Joyce 2005)—or parameters for which only marginally sufficient data are available for their estimation. In Bayesian analyses, when the data do not strongly inform a parameter estimate, the prior on that parameter can have a large effect.

In the first of these 2 tests, I used simulated data sets to determine if among-partition rate variation (APRV) is necessary for the long-tree problem. Using Seq-Gen, I simulated ten 2400 bp, 3-partition data sets using the same model and tree as in the previous simulated data analyses but under each of 4 partition rate treatments, "high APRV," "low APRV," "no APRV," and "no APRV control" (see Table 5). In the first 3 of these treatments, overall TL was held constant and the partition rate multipliers were manipulated to yield decreasing contrast between the fastest- and slowest evolving data partitions. The fourth treatment consisted of data sets simulated under equal partition rates but with a TL equal to that of the slowest data partition in the high-APRV treatment. Note that, in the no-APRV-control treatment, the data were still simulated and analyzed using a partitioned framework. MrBayes v3.1 analyses were observed over 2 million generations and scored for whether they had found the correct short-tree solution in that time.

The results of the test showed that partition rate variation is not necessary for the long-tree problem to occur. Both the high-APRV and the no-APRV-control treatments showed a tendency to become trapped in the region of long-tree solutions (Table 6). What the data sets of these 2 treatments have in common is the existence of 1 or more data partitions that contain few variable sites, a situation that is likely to reduce the identifiability of the parameters for those partitions and increase the control of the prior over the final estimate. As expected in this case, the TL estimates suggest a strong effect of the branch length prior, with the 10 runs averaging TL = 4.4, just a bit shorter than the prior TL of 4.7 (47 internal branches multiplied by 0.1). Furthermore, in additional runs not shown here, I found that analyses of the *Kikihia* empirical data set consistently found the short-tree solution when the 2 lowest-data partitions (first- and second-codon positions) were combined.

TABLE 6. Effect of APRV on the tendency of MrBayes analyses to become stuck on long-tree solutions

	No APRV	Low APRV	High APRV	No APRV control
Number of correct solutions	10	10	6	0
Number of long-tree solutions	0	0	4	10

Note: In the no-APRV-control treatment, all data partitions were simulated under the parameter values used for the slowest partition of the "high-APRV" treatment.

I found additional indications that marginal identifiability of 1 or more parameters is involved while exploring the effects of the number of branches in the tree. When I reduced the number of taxa in the cicada mtDNA data set to 32, by removing taxon duplicates (most of which are nearly identical to other sequences in the data set), the resulting analyses were less likely to become stuck on the long-tree solution (results not shown). Removing taxon duplicates reduced the number of branches in the tree but had little effect on the overall TL or information content of the data set. This situation recalls the results of Marshall et al. (2006, p. 998), who briefly noted long-tree solutions in partitioned analyses and found that placing a stronger prior on short branches sometimes caused MCMC chains to avoid the long-tree trap.

The appropriateness of the substitution models employed in these analyses, as well as that of the partitioning schemes, was confirmed using some of the most commonly employed model selection routines. Furthermore, the Bayesian analyses all recovered reasonable estimates of the parameter values when the MCMC chains fortuitously landed on the short-tree solutions, so the parameters were apparently not strictly nonidentifiable. Apparently, commonly used model-testing procedures can lead the investigator to select a model containing only marginally identifiable parameters, and when this is done in the context of a partitioned Bayesian analysis, the alternative long-tree solution becomes a problem. If so, the patterns described here confirm the concerns of Rannala (2002, p. 759), who warned that "overly complex models with many nonidentifiable parameters may lead to large correlations among parameters in the posterior density, possibly retarding convergence of an MCMC algorithm." The findings here may indicate that MCMC chains can have serious problems with convergence even as parameters approach nonidentifiability.

OTHER MEANS OF GUIDING ANALYSES TO IMPROVED BRANCH LENGTH SOLUTIONS

MrBayes offers detailed control over prior probability distributions for all model parameters as well as settings that determine the frequency and nature of proposed changes to parameter values during MCMC (see the props command). In addition, the number of

concurrently run chains can be increased or decreased, and chain "heating" values can be altered to improve the odds of escaping local maxima. There are probably many means of averting the long-tree syndrome described here, and I now mention just 2 that I have encountered during experimentation in addition to the short-starting-tree solution (see Brown et al. 2010). First, as mentioned earlier, many analyses of simulated and empirical data sets appear less likely to become trapped on local TL optima as the mean branch length prior is reduced below the MrBayes default mean of 0.1. This was observed in the study of Marshall et al. (2006). Second, the prior on the partition rate multiplier, which defaults to a dirichlet(1,1,1) distribution, can be adjusted to place a greater probability on a specific order of relative partition rates. For example, setting the prior to dirichlet(1,1,100) places greater emphasis on comparatively large third partition *m* values and usually prevents the correlated parameter shifts that commit some analyses to long-tree solutions. However, some investigators may not be satisfied with the manipulation of priors for these purposes—more informative priors have stronger effects on posterior parameter estimates. The number of chains may be increased, but because the rest of the chains may remain lost (as demonstrated above), the effectiveness of the overall tree search may be decreased. It may be that when accurate branch length estimates are required in difficult cases, it is sometimes necessary to turn to maximum-likelihood (ML) methods over Bayesian ones to avoid the influence of the prior on parameter estimates.

Examples of Published Partitioned Bayesian Analyses with TL Problems

To gauge the frequency of the long-tree problem in the published phylogenetic literature, I conducted an informal survey of Internet-accessible files using Google Scholar (scholar.google.com) and a keyword set consisting of "partition," "MrBayes," "mitochondrial," "codon," "phylogeography," and "TreeBASE." The latter 2 terms were intended to bias the sample toward studies with larger numbers of taxa and easily accessed data sets. I examined the first 24 studies that fit these criteria; many proved unusable here because the data sets were not accessible, the character sets were not clearly specified in the TreeBASE files (http://www.treebase.org/treebase/index.html), or the partitioning details (e.g., parameter linking) were not entirely specified. For the remaining studies, I compared the results of 2 analysis methods, the first of which replicated the methodology originally described by the authors as closely as possible (the "published method") and the second of which placed a shorter prior on the mean branch length ("short branch method"—usually brlens = exponential:unconstrained[100]). Each analysis consisted of a pair of MC^3 runs begun from different random starting trees. See supplementary Appendix 1 for more details.

Analyses from 6 data sets from the initial list of 24 papers exhibited strong distortions of the final TL and rate multipliers (Table 7). In all cases, application of the shorter branch length prior improved the ln *L* score, the relative partition rates, or both. Three data sets, from the Bryson et al. (2007), Leaché and Mulcahy (2007), and McGuire et al. (2007) studies, behaved similarly to the *Kikihia* empirical data set above. Analyses using the published method (the default MrBayes branch length prior) yielded very long trees and nonsensical relative partition rates. Analyses under a shorter branch length prior dramatically shortened the TL, improved the log likelihood, and returned empirically sensible rate multipliers in 2 of these 3 examples. In the McGuire et al. (2007) case, the rate exaggeration was reduced by the shorter branch length prior but the relative rates remained incorrect (i.e., second-codon position still fastest). The remaining 3 cases (Uit de Weerd et al. 2004; Lymberakis et al. 2007; Hunter and Halanych 2008) showed a related but different pattern. In these cases, the rate ordering of the partitions appeared reasonable, but the contrast between the rate categories was much greater than expected given other empirical studies (e.g., Lin and Danforth 2004). In the Lymberakis et al. study, for example, the CYTb data partition was estimated to be evolving at 40 times the rate of the 16S data partition. Shortening the branch length prior in 2 of these cases shortened the final TL estimate, reduced the contrast in relative partition rates, dramatically improved the likelihood score, and left the relative order of the partition rates unchanged. Interestingly, in the Uit de Weerd case, shortening the branch length prior reduced the apparent distortion of the relative partition rates, but at the expense of the final likelihood score (Table 7).

As in the analyses of the *Kikihia* data set, some of the published method analyses yielded different results between the paired MC^3 runs (similar to Table 2, Analysis VI), indicating dependence of the final outcome on the initial random starting conditions. In the Lymberakis et al. trials, 2 of the published method analyses found solutions closely resembling those illustrated in Table 7, whereas in the third published method replicate just one of the paired runs found the short-tree solution identified under the short branch method. During the Hunter and Halanych trials, 2 of the 3 published method analyses found different solutions between the paired runs. In these cases, both the PSRF and the ASDSF statistics indicated clear problems (e.g., a final PSRF of 1.230 for TL and a final ASDSF of 0.036288 for one of the Hunter and Halanych analyses; see also Analysis VI, Table 2). The rest of the published method replicates resulted in the same solution for both paired runs, as did all the analyses using the short branch method.

In a separate set of trials (results not shown), I ran analyses of the 6 published data sets using short starting trees, rather than short branch length priors, as discussed earlier. One of these analyses (of the Bryson et al. dataset) consistently found the short-tree solution when a short starting tree was used, but the other 5 were unaffected. This, together with the interesting result from the

TABLE 7. Partitioned MrBayes analyses of published data sets demonstrating likely failure to accurately estimate tree length (TL) and partition rate multipliers

Lymberakis et al. (2007)

Rana spp. frogs, 76 sequences, 1055 bp

Published method
Final ASDSF = 0.004988
Largest bipartition PSRF = 1.000

Parameter	Run 1	Run 2	PSRF
Ln *L*	−5532.9	−5535.3	
TL	11.057	11.02	1.000
m_{CYTb}	2.004	2.004	1.000
m_{16S}	0.050	0.050	1.000

Short branch method
Final ASDSF = 0.007573
Largest bipartition PSRF = 1.014

Parameter	Run 1	Run 2	PSRF
Ln *L*	−5388.2	−5386.6	
TL	1.184	1.180	1.001
m_{CYTb}	1.544	1.543	1.000
m_{16S}	0.485	0.486	1.000

Bryson et al. (2007)

Lampropeltis spp. snakes, 34 sequences, 868 bp

Published method
Final ASDSF = 0.002820
Largest bipartition PSRF = 1.006

Parameter	Run 1	Run 2	PSRF
Ln *L*	−4290.8	−4290.4	
TL	6.135	6.094	1.001
$m_{ND4\ 1}$	0.100	0.099	1.000
$m_{ND4\ 2}$	3.067	3.064	1.000
$m_{ND4\ 3}$	0.500	0.504	1.001
m_{tRNA}	0.098	0.099	1.001

Short branch method
Final ASDSF = 0.001999
Largest bipartition PSRF = 1.003

Parameter	Run 1	Run 2	PSRF
Ln *L*	−4277.7	−4275.2	
TL	0.992	0.994	1.000
$m_{ND4\ 1}$	0.511	0.511	1.000
$m_{ND4\ 2}$	0.154	0.154	1.000
$m_{ND4\ 3}$	2.686	2.687	1.000
m_{tRNA}	0.514	0.512	1.000

Mcguire et al. (2007)

Crotaphytidae lizards, 407 sequences, 1707 bp

Published method
Final ASDSF = 0.008565
Largest bipartition PSRF = 1.134

Parameter	Run 1	Run 2	PSRF
Ln *L*	−19112	−19115	
TL	16.047	16.181	1.022
$m_{ND2\ 1}$	0.092	0.092	1.001
$m_{ND2\ 2}$	0.038	0.039	1.030
$m_{ND2\ 3}$	0.311	0.309	1.010
m_{tRNA}	0.124	0.012	1.009
$m_{CYTb\ 1}$	0.051	0.049	1.017
$m_{CYTb\ 2}$	10.331	10.351	1.023
$m_{CYTb\ 3}$	0.393	0.387	1.021

Short branch method
Final ASDSF = 0.009719
Largest bipartition PSRF = 1.164

Parameter	Run 1	Run 2	PSRF
Ln *L*	−19103	−19103	
TL	8.039	8.007	1.006
$m_{ND2\ 1}$	0.188	0.187	1.000
$m_{ND2\ 2}$	0.078	0.078	1.000
$m_{ND2\ 3}$	0.624	0.623	1.000
m_{tRNA}	0.248	0.241	1.006
$m_{CYTb\ 1}$	0.103	0.102	1.001
$m_{CYTb\ 2}$	8.560	8.575	1.003
$m_{CYTb\ 3}$	0.789	0.785	1.002

Hunter and Halanych (2008)

Astrotoma agassizii brittle star, 65 sequences, 985 bp

Published method
Final ASDSF = 0.006456
Largest bipartition PSRF = 1.011

Parameter	Run 1	Run 2	PSRF
Ln *L*	−2858.4	−2860.4	
TL	11.629	11.658	1.000
m_{16S}	2.075	2.075	1.000
m_{CO2}	0.031	0.030	1.000

Short branch method
Final ASDSF = 0.006309
Largest bipartition PSRF = 1.014

Parameter	Run 1	Run 2	PSRF
Ln *L*	−2819.1	−2822.8	
TL	0.997	0.989	1.002
m_{16S}	1.682	1.679	1.001
m_{CO2}	0.385	0.388	1.001

Leache and Mulcahy (2008)

Sceloporus spp. lizards, 123 sequences, 1606 bp.

Published method
Final ASDSF = 0.010014
Largest bipartition PSRF = 1.014

Parameter	Run 1	Run 2	PSRF
Ln *L*	−11830	−11827	
TL	23.108	23.144	1.000
m_{tRNA}	0.0457	0.0456	1.000
$m_{ND4\ 1}$	6.351	6.353	1.001
$m_{ND4\ 2}$	0.0181	0.018	1.000
$m_{ND4\ 3}$	0.323	0.321	1.001

Short branch method
Final ASDSF = 0.005542
Largest bipartition PSRF = 1.016

Parameter	Run 1	Run 2	PSRF
Ln *L*	−11746	−11751	
TL	2.062	2.067	1.002
m_{tRNA}	0.541	0.543	1.001
$m_{ND4\ 1}$	0.776	0.776	1.000
$m_{ND4\ 2}$	0.209	0.208	1.000
$m_{ND4\ 3}$	3.789	3.783	1.001

Uit de Weerd et al. (2004)

Alopiinae snails, 53 sequences, 1794 bp

Published method
Final ASDSF = 0.003542
Largest bipartition PSRF = 1.005

Parameter	Run 1	Run 2	PSRF
Ln *L*	−19330	−19330	
TL	13.565	13.552	1.000
$m_{CO1\ 1}$	0.177	0.178	1.000
$m_{CO1\ 2}$	0.011	0.011	1.000
$m_{CO1\ 3}$	7.470	7.465	1.002
m_{12S}	0.239	0.241	1.002
m_{ITS1}	0.050	0.050	1.000
m_{ITS2}	0.080	0.081	1.002

Short branch method
Final ASDSF = 0.004828
Largest bipartition PSRF = 1.072

Parameter	Run 1	Run 2	PSRF
Ln *L*	−19502	−19509	
TL	3.619	3.609	1.000
$m_{CO1\ 1}$	0.430	0.432	1.000
$m_{CO1\ 2}$	0.037	0.037	1.003
$m_{COI\ 3}$	6.042	6.036	1.000
m_{12S}	0.644	0.646	1.000
m_{ITS1}	0.164	0.164	1.000
m_{ITS2}	0.254	0.253	1.002

Notes: "Published method" indicates results obtained using the methods reported by the authors. "Short branch method" indicates results obtained under shorter branch length priors. For each run (nruns = 2) of each analysis, the estimated TL and partition rate multipliers (m_i) are shown along with the final harmonic mean log likelihood, except for McGuire example where the arithmetic mean log likelihood is shown (both were flagged as uncertain by MrBayes due to exclusion of extreme values). The PSRF comparing the 2 parameter estimates within each analysis is given, along with the maximum PSRF observed for the bipartition posterior probabilities. The ASDSF observed at the end of the analysis is also shown. Each published or short branch analysis was replicated 3 or 4 times, with one of the results shown here in each case—the remaining results are summarized in the text. Shading is intended only to visually offset results from different studies. tRNA = transfer RNA.

Uit de Weerd data set, in which improved log-likelihood scores were associated with more extremely distorted rate multipliers, suggests that different causes of the TL problem are involved in different cases.

Significantly, in almost every analysis shown in Table 7, there is no way to tell from the convergence diagnostics that the published method analyses may have problems. Even in the 2 cases in which the original rate multiplier estimates are clearly wrong, the ASDSF and PSRF statistics are all but identical between the published and short branch analyses. The only sure way for an investigator to recognize that something is amiss is by examining the posterior parameter estimates in light of prior expectations.

The published papers in 4 of the 6 cases examined in detail contain clear signs that the analyses suffered from the long-tree problem. The scale bar of the tree in figure 2, page 121, of Lymberakis et al. indicates a pairwise genetic distance of at least 1.5 substitutions/site between their "Far Eastern clade" and the remaining ingroup sequences. In contrast, their table of genetic distances (p. 120) shows corrected pairwise divergences of 0.21–0.24 for these relationships, a disparity that matches the difference in the TLs estimated by the published method and short branch analyses in Table 7. A similar contradiction is apparent between the Bayesian consensus phylogram of Bryson et al. (2007) and the pairwise distance estimates given in their table 3 (p. 678), with the contradiction again matching the difference in the 2 solutions found in Table 7 here. In both studies, the pairwise distances were apparently calculated in Paup* and were therefore unaffected by the problems with the Bayesian analyses. Although the other papers do not present tables of independently calculated genetic distances, other signs of the problem do appear. Leaché and Mulcahy (2007) estimated divergence times for the clades in their tree; when combined with the genetic distances implied by the scale bar on their Bayesian phylogram, these divergence times suggest per-lineage mtDNA evolutionary rates of approximately 0.07 substitutions/site/million years, several times greater than rates estimated by other studies (e.g., Pereira and Baker 2006). Similar contradictions are apparent between scale bars on phylograms and other genetic data presented in Hunter and Halanych (2008). Uit de Weerd et al. (2004) presented only a cladogram, and McGuire et al. (2007) presented a phylogram but with no scale information, so that there is no way to know from reading the papers whether or not the analyses suffered from the branch length problem.

Significance of the Long-Tree Problem

Attraction of MCMC chains to long-tree solutions is a concern for studies at a variety of evolutionary levels. Studies of closely related taxa, especially intrageneric studies, will often involve data partitions with few variable sites and data sets with short ML branch length values. But deeper level partitioned data studies can exhibit the problem as well when slowly evolving markers are used, and because the ML branch length depends on the number of taxa sampled as well as the maximum genetic divergence of the group being studied.

So far it appears that the principal problem caused by the long-tree phenomenon is gross misestimation of evolutionary divergence, which has clear implications for studies that use genetic divergence values to date molecular phylogenetic trees or to estimate molecular clock rates. Although many molecular clock studies use pairwise genetic divergence values obtained from programs such as Paup* (Swofford 1998), pairwise estimation methods tend to underestimate real divergence values due to limitations of the evolutionary models implemented. In contrast, as taxon sampling increases, phylogenetic distances tend to increase as previously missed homoplasy is "unmasked" (the "node-density artifact"—Fitch and Beintema 1990; Webster et al. 2003; Venditti et al. 2006). Therefore, future refinements of our understanding of genetic divergence may depend on estimates from well-sampled phylogenetic trees.

The implications of the long-tree problem for topological inference are not yet clear—I have not carefully explored the effects of the long-tree problem on branch support. It is striking that both the topology and the relative branch lengths appear so similar between the 2 analyses shown in Figure 1. One node in the *Kikihia* phylogeny is consistently estimated with a roughly 30% higher posterior probability in long-tree solutions (the 89% node of Analysis II, Fig. 1b)—it would be interesting to investigate whether this is related to recent findings that long internal branch priors can bias posterior probabilities upward (Yang and Rannala 2005). It is difficult to believe that the severe distortions of among-site rate variation parameters observed in this study can occur generally with no effect on topological inferences based on characters in the affected data partitions. Two other published studies have demonstrated topological effects of failure to correctly accommodate partition rate variation, a related problem that leads to distortions of rate variation parameters similar to those observed here (Marshall et al. 2006; Angelini and Jockusch 2007). Lack of obvious topological effects in the data sets examined here might be due to the small number of phylogenetically informative sites present in the affected data partitions, in which case the most seriously affected studies might be those with a large fraction of the informative sites scatted across multiple low-data partitions. This hypothesis could be investigated with simulated data sets.

Supplementary Material

Supplementary material can be found at: http://www.sysbio.oxfordjournals.org/.

Funding

National Science Foundation (DEB-0089946, DEB-0422386, DEB-0529679, and DEB-0720664 to Chris Simon).

Acknowledgments

John Cooley, Kathy Hill, Elizabeth Jockusch, Paul Lewis, and Chris Simon offered expertise, constructive criticism, and software that substantially influenced the findings of this research. Robert W. Bryson, Jr. and Frank Burbrink kindly contributed analysis details. Two editors and 2 external reviewer provided extensive suggestions that considerably improved the paper. Extensive computational resources, without which this study could not have been attempted, were provided by the Bioinformatics Facility of the Biotechnology/Bioservices Center, University of Connecticut, Storrs, Connecticut, and by the Bioportal at the University of Oslo, Norway (http://www.bioportal.uio.no).

References

Angelini D.R., Jockusch E.L. 2007. Relationships among pest flour beetles of the genus *Tribolium* (Tenebrionidae) inferred from multiple molecular markers. Mol. Phylogenet. Evol. 46:127–141.

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

Brown J.M., Hedtke S.M., Lemmon A.R., Lemmon E.M. Forthcoming 2010. When trees grow too long: investigating the causes of highly inaccurate Bayesian branch length estimates. Syst. Biol. 59.

Brown J.M., Lemmon A.R. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56:643–655.

Bryson R.W. Jr., Pastorini J., Burbrink F.T., Forstner M.R.J. 2007. A phylogeny of the *Lampropeltis mexicana* complex (Serpentes: Colubridae) based on mitochondrial DNA sequences suggests evidence for species-level polyphyly within Lampropeltis. Mol. Phylogenet. Evol. 43:674–684.

Fitch W.M., Beintema J.J. 1990. Correcting parsimonious trees for unseen nucleotide substitutions: the effect of dense branching as exemplified by ribonuclease. Mol. Biol. Evol. 7:438–443.

Gelman A., Rubin D.B. 1992. Inference from iterative simulation using multiple sequences (with discussion). Stat. Sci. 7:457–511.

Goloboff P.A., Pol D. 2005. Parsimony and Bayesian phylogenetics. In: Albert V.A., editor. Parsimony, phylogeny, and genomics. Oxford: Oxford University Press. p. 148–159.

Huelsenbeck J.P., Ronquist F. 2005. Bayesian analysis of molecular evolution using MrBayes. In: Nielsen R., editor. Statistical methods in molecular evolution. New York: Springer.

Hunter R.L., Halanych K.M. 2008. Evaluating connectivity in the brooding brittle star *Astrotoma agassizii* across the Drake Passage in the Southern Ocean. J. Hered. 99:137–148.

Jukes T.H., Cantor C.R. 1969. Evolution of protein molecules. In: Munro H.N., editor. Mammalian protein metabolism. New York: Academic Press. p. 21–132.

Kass R.E., Raftery A.E. 1995. Bayes factors. J. Am. Stat. Assoc. 90:733–795.

Leaché A.D., Mulcahy D.G. 2007. Phylogeny, divergence times and species limits of spiny lizards (*Sceloporus magister* species group) in western North American deserts and Baja California. Mol. Ecol. 16:5216–5233.

Lin C.-P., Danforth B.N. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol. Phylogenet. Evol. 30: 686–702.

Lymberakis P., Poulakakis N., Manthalou G., Isigenopolous C.S., Magoulas A., Mylonas M. 2007. Mitochondrial phylogeography of *Rana* (*Pelophylax*) populations in the Eastern Mediterranean region. Mol. Phylogenet. Evol. 44:115–125.

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

Marshall D.C., Slon K., Cooley J.R., Hill K.B.R., Simon C. 2008. Steady Plio-Pleistocene diversification and a 2-million-year sympatry threshold in a New Zealand cicada radiation. Mol. Phylogenet. Evol. 48:1054–1066.

McGuire J.A., Linkhem C.W., Koo M.S., Hutchison D.W., Lappin A.K., Orange D.I., Lemos-Espinal J., Riddle B.R., Jaeger J.R. 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. Evolution. 61:2879–2897.

Miya M., Satoh T.P., Nishida M. 2005. The phylogenetic position of toadfishes (order Batrachoidiformes) in the higher rayfinned fish as inferred from partitioned Bayesian analysis of 102 whole mitochondrial genome sequences. Biol. J. Linn. Soc. 85: 289–306.

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

Pereira S.L., Baker A.J. 2006. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Mol. Biol. Evol. 23:1731–1740.

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

Rambaut A. 2002. Phyl-O-Gen: phylogenetic tree simulator package. Version 1.1. Oxford: University of Oxford.

Rambaut A., Drummond A.J. 2003. Tracer v1.3. Oxford: University of Oxford. Available from: http://evolve.zoo.ox.ac.uk/.

Rambaut A., Grassly N.C. 1997. Seq-Gen: an application for the Monte-Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13:235–238.

Rannala B. 2002. Identifiability of parameters in MCMC Bayesian inference of phylogeny. Syst. Biol. 51:754–760.

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

Ronquist F., Huelsenbeck J.P., van der Mark P. 2005. MrBayes 3.1 Manual. Version 3.1. Available from: http://mrbayes.csit.fsu.edu/manual.php.

Soltis D.E., Gitzendanner M.A., Soltis P.E. 2007. A 567-taxon data set for angiosperms: the challenges posed by Bayesian analyses of large data sets. Int. J. Plant Sci. 168:137–157.

Sullivan J., Joyce P. 2005. Model selection and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 36:445–466.

Sullivan J., Swofford D.L., Naylor G.J.P. 1999. The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. Mol. Biol. Evol. 16:1347–1356.

Swofford D.L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Tierney L. 1994. Markov chains for exploring posterior probability distributions. Ann. Stat. 22:1701–1728.

Uit de Weerd D.R., Piel W.H., Gittenberger E. 2004. Widespread polyphyly among Alopiinae snail genera: when phylogeny mirrors biogeography more closely than morphology. Mol. Phylogenet. Evol. 33:533–548.

Venditti C., Meade A., Pagel M. 2006. Detecting the node-density artifact in phylogeny reconstruction. Syst. Biol. 55:637–643.

Webster A.J., Payne R.J., Pagel M. 2003. Molecular phylogenies link rates of evolution and speciation. Science. 301:478.

Yang Z. 1994a. Estimating the pattern of nucleotide substitution. J. Mol. Evol. 39:105–111.

Yang Z. 1994b. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306–314.

Yang Z., Rannala B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. Syst. Biol. 54:455–470.