*Syst. Biol.* 50(1):67–86, 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

THOMAS R. BUCKLEY,[1] CHRIS SIMON,[1,2] AND GEOFFREY K. CHAMBERS[1]

[1]*Institute for Molecular Systematics, School of Biological Sciences, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand; E-mail: tbuckley@duke.edu, geoff.chambers@vuw.ac.nz*

[2]*Department of Ecology and Evolutionary Biology, University of Connecticut, U-43 Storrs, CT, USA; E-mail: csimon@uconnvm.uconn.edu*

*Abstract*.—We have investigated the effects of different among-site rate variation models on the estimation of substitution model parameters, branch lengths, topology, and bootstrap proportions under minimum evolution (ME) and maximum likelihood (ML). Specifically, we examined equal rates, invariable sites, gamma-distributed rates, and site-specific rates (SSR) models, using mitochondrial DNA sequence data from three protein-coding genes and one tRNA gene from species of the New Zealand cicada genus *Maoricicada*. Estimates of topology were relatively insensitive to the substitution model used; however, estimates of bootstrap support, branch lengths, and R-matrices (underlying relative substitution rate matrix) were strongly influenced by the assumptions of the substitution model. We identified one situation where ME and ML tree building became inaccurate when implemented with an inappropriate among-site rate variation model. Despite the fact the SSR models often have a better fit to the data than do invariable sites and gamma rates models, SSR models have some serious weaknesses. First, SSR rate parameters are not comparable across data sets, unlike the proportion of invariable sites or the alpha shape parameter of the gamma distribution. Second, the extreme among-site rate variation within codon positions is problematic for SSR models, which explicitly assume rate homogeneity within each rate class. Third, the SSR models appear to give severe underestimates of R-matrices and branch lengths relative to invariable sites and gamma rates models in this example. We recommend performing phylogenetic analyses under a range of substitution models to test the effects of model assumptions not only on estimates of topology but also on estimates of branch length and nodal support. [Among-site rate variation; bootstrapping; insect mitochondrial DNA; likelihood ratio test; *Maoricicada*; maximum likelihood; nucleotide substitution models.]

The increasing use of maximum likelihood (ML) methods in phylogenetic and molecular evolutionary studies is largely a result of its desirable and well-understood statistical properties (Wald, 1949; Edwards, 1992; Gaut and Lewis, 1995; Rogers, 1997), the ease with which competing hypotheses can be evaluated (Kishino and Hasegawa, 1989; Huelsenbeck and Rannala, 1997), and the development of explicit and biologically realistic models of nucleotide substitution (e.g., Yang, 1993, 1994a; Galtier and Gouy, 1998). In addition, the implementation of ML methods in user-friendly software packages (e.g., PHYLIP, Felsenstein, 1993; PUZZLE, Strimmer and Von Haesseler, 1996; and PAUP*4.0, Swofford, 1998) have made ML methods accessible to the wider systematics community. Continuing advances in computer processing speed and algorithm implementation (e.g., Lewis, 1998; Swofford, 1998) have to an extent alleviated some problems associated with the intensive computational burden of likelihood calculations. Several simulation studies have shown that the ML optimality criterion tends to be an accurate estimator of phylogeny over a wider area of tree space—that is, over different combinations of branch lengths—than many other methods (e.g., Hillis et al., 1994; Huelsenbeck, 1995a, 1995b; Yang, 1996a). The accuracy of ML depends largely on a lack of systematic error, that is, a good fit between the assumptions of the substitution model and the true underlying evolutionary process (Swofford et al., 1996; Rogers, 1997). ML may become inconsistent if an overly simplistic substitution model is used to select an optimal tree (Nei, 1991; Swofford et al., 1996), potentially leading to strong statistical support for incorrect phylogenetic hypotheses (e.g., Lockhart et al., 1996; Sullivan and Swofford, 1997). This observation has stimulated much interest in the assumptions made by the various substitution models required for likelihood analyses (e.g., Yang et al., 1994, 1995). Those models do attempt to account for the major features of nucleotide sequence

evolution such as unequal base composition among sites (Felsenstein, 1981), transition bias (Kimura, 1980), and among-site rate variation (e.g., Yang, 1993).

Among-site rate variation is one ubiquitous property of sequence evolution that, where ignored, can drastically affect the estimation of topology (Sullivan et al., 1995; Lockhart et al., 1996), branch lengths (Waddell and Steel, 1997), substitution model parameters (Wakeley, 1994, 1996), and bootstrap proportions (Frati et al., 1997; Sullivan et al., 1997) under all optimality criteria (Kuhner and Felsenstein, 1994). Ignoring among-site rate variation also can seriously affect the power of likelihood ratio tests (LRTs; Huelsenbeck et al., 1997; Zhang, 1999) and the estimation of *P*-values in the Kishino and Hasegawa (1989) test (Waddell et al., 1999; Buckley et al., unpubl. mss.) and multiple comparison tests (Shimodaira and Hasegawa, 1999; Buckley et al., unpubl.). The detrimental effect of ignoring among-site rate variation on the estimation of topology has been demonstrated in both empirical studies (Sullivan et al., 1995; Lockhart et al., 1996; Takezaki and Gojobori, 1999) and simulations (Kuhner and Felsenstein, 1994; Tateno et al., 1994; Yang, 1996a). Genes that exhibit extreme among-site rate variation have less phylogenetic information than a gene of the same length with moderate or no among-site rate variation. This is because fewer sites are free to vary and those that do vary may be evolving at a high rate; thus many similarities among the tips of a tree will be due to homoplasy and not homology (Sullivan et al., 1995). The problem of among-site rate variation becomes particularly acute when rates of change vary among lineages (Cunningham et al., 1998). In this situation, accurate optimization of branch lengths is more critical if phylogenetic analysis is to be accurate. Waddell and Steel (1997) and Waddell et al. (1997) noted that different among-site rate variation models will often converge on similar topologies despite their variations in estimates of branch lengths and transition/transversion ratios (TS:TV). Because of the nonlinear effect of among-site rate variation on inferred branch lengths, differences between estimates determined with different models tend to be most pronounced for long branches in the tree (Waddell et al., 1997). Thus, to obtain accurate branch length estimates, especially for deeper divergences, biologically realistic models should be used. Reliable branch length estimates are especially important for dating internal nodes with a molecular clock (Yang, 1996b; Waddell and Steel, 1997; Rambaut and Bromham, 1998).

We have used mitochondrial DNA sequences from the cytochrome oxidase subunit I (COI), ATPase subunits 6 (A6) and 8 (A8), and the transfer RNA aspartic acid ($tRNA^{Asp}$) gene to reconstruct phylogenetic relationships among species of the genus *Maoricicada* (Dugdale, 1972), a group of predominantly montane New Zealand cicadas (Fleming, 1971; Dugdale and Fleming, 1978). The biological motivation for selection of this taxonomic group was to study the origin and diversification of the New Zealand alpine biota by using *Maoricicada* as a model taxon. The evolutionary and biogeographic implications of the phylogenetic results presented here are discussed elsewhere (Buckley et al., 2001). The mitochondrial genes analyzed here are evolving under a range of constraints and represent typical markers used in many molecular systematic studies.

We contrasted several methods for modeling among-site rate variation and examined their effects on estimates of topology, branch lengths, bootstrap values, and substitution model parameters under the minimum evolution (ME) and ML optimality criteria. We demonstrate that the manner in which among-site rate variation is accounted for has marked effects on several aspects of phylogenetic analysis. In particular, estimates of nodal bootstrap support are highly dependent on the among-site rate variation model used. We also identify a possible example where ML and ME are behaving in an inaccurate manner because of the presence of long adjacent branches and the use of a biologically unrealistic and thus inappropriate among-site rate variation model (see Waddell et al., 1999). We discuss and evaluate the underlying biological assumptions of a range of among-site rate variation models and their appropriateness for analyzing the data presented here.

## METHODS

### *Species Sampling and Laboratory Protocols*

We sampled all described species and subspecies of *Maoricicada* except *M. otagoensis*

*maceweni*; locality information will be presented elsewhere (Buckley et al., 2001). At least two individuals were sequenced for each species or subspecies, on different days, to guard against possible contamination or specimen mix-up and to search for possible cryptic species. Whole genomic DNA was extracted by using the salting-out protocol described by Sunnucks and Hales (1996). We amplified two mitochondrial DNA targets. The first, 819 bp from the COI gene, was amplified by using the primers C1-J-2195 and TL2-N-3014 (Simon et al., 1994). The second, a 771-bp region from the $tRNA^{Asp}$, $tRNA^{Leu}$, A8, and A6 genes, was amplified by using the primer pair TK-J-3799 (GGCTGAAAGTAA GTAATGGTCTCT) and A6-N-4570 (AAG ACTGAATTATACAAACGGCTA). All specimens were collected from the field except for two *M. iolanthe* sequences, which were obtained from museum specimens collected in 1971. For both *M. iolanthe* individuals, target mtDNA gene regions were amplified in a series of overlapping fragments; primer sequences will be given elsewhere (Buckley et al., 2001). Polymerase chain reaction (PCR) products were gel-purified and cycle-sequenced with the Perkin-Elmer Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit, according to the manufacturer's instructions. Cycle-sequencing products were purified by ethanol precipitation and analyzed by electrophoresis on an ABI Prism™ 377 DNA Sequencer. Sequences were manually aligned by using ESEE3.2 (Cabot and Beckenbach, 1989), facilitated by the conserved amino acid sequence and lack of indels.

### *Patterns of Variation and Substitution Model Selection*

We calculated the number of varied sites (sites observed to vary), parsimony-informative sites, and the base frequencies for each gene and codon position. Shifts in base composition among taxa were examined by using $\chi^2$ heterogeneity tests as implemented in PAUP*4.0b2a (Swofford, 1998). Deviations from base composition stationarity were tested on all sites and on parsimony sites only, to assess the potentially confounding effect of unvaried sites (sites observed to be constant), which by definition have stationary base frequencies (Waddell et al., 1999). We also tested the stationarity assumption on each of the four gene coding regions and each of the three codon positions. Using MEGA (Kumar et al., 1993), we calculated the number of twofold and fourfold degenerate sites in the first and third positions of each protein-coding gene.

Phylogenetic analyses were conducted under the ML (Felsenstein, 1981), ME (Kidd and Sgaramella-Zonta, 1971; Rzhetsky and Nei, 1992), and maximum parsimony (MP; Fitch, 1971) optimality criteria as implemented in PAUP*. Initially, a heuristic MP tree search was performed with tree-bisection-reconnection (TBR) branch swapping under equal weights. This search converged on five equally parsimonious trees. The likelihood of one of the five MP trees was then calculated using several substitution models, according to the methods of Frati et al. (1997) and Sullivan et al. (1997). Results from the other four most-parsimonious trees were essentially identical, as might be expected from results of other studies (see Sullivan et al., 1996, 1997). The substitution models tested are those of Jukes and Cantor (1969; JC69), Kimura (1980; K80), Hasegawa et al. (1985; HKY85), and the general-time reversible model (GTR; e.g., Yang, 1994a). We also accommodated among-site rate variation by using six categories of rate heterogeneity models. The first category assumed that a proportion of sites are invariable (e.g., I; Hasegawa et al., 1985); the second category assumed that all sites are free to vary, with rates among sites following a discrete approximation to the gamma distribution ($\Gamma$ Yang, 1994b); and the third category assumed that a proportion of sites is invariable and the remainder are free to vary following a gamma distribution ($I + \Gamma$; Gu et al., 1995). For the gamma distribution we used eight rate categories to avoid underestimating the $\alpha$ shape parameter, as recommended by Yang (1994b) for cases where among-site rate variation is extreme.

For the fourth category, we partitioned the characters into first, second, and third codon positions and all $tRNA^{Asp}$ sites and then estimated the gamma parameter separately for each of these four partitions after Yang (1996c), using PAML2.0 (Yang, 1997). We refer to this model as the $\Gamma_4$ model. Base frequencies and underlying relative substitution rate matrices (R-matrices) were assumed

to be constant across the four character partitions. Although optimizing base frequency and substitution type parameters separately for each of the four partitions is more realistic, we have kept these parameters homogeneous because we wanted to examine the effects of the assumption that the distribution of among-site rate variation differs among the partitions in isolation from other factors.

The fifth and sixth models are site-specific rate (SSR) models, in which we identified rate classes a priori according to the functional properties of a site (e.g., Swofford et al., 1996). We identified 10 potential rate classes (referred to as the $SSR_{10}$ model): the three codon positions for each of the three protein-coding genes, and the set of all sites in the $tRNA^{Asp}$ gene. We also pooled each of the codon positions across genes, which yielded four rate classes: each codon position plus the $tRNA^{Asp}$ sites (referred to as the $SSR_4$ model). Rate classes were specified by using the site rates/rate sets option in PAUP*. For all SSR models, the relative rate for each site was assumed to be equal within a class and these rates were optimized by using ML. The $GTR + SSR_{10}$ model is the most general of the site-specific rate models.

The likelihoods of the MP topology, calculated under the above range of substitution models, were evaluated by using LRTs (Goldman, 1993; Frati et al., 1997; Huelsenbeck and Crandall, 1997). The LRT statistic ($\delta$) is defined as: $\delta = -2(\ln L_1 - \ln L_0)$, where $\ln L_0$ is the natural logarithm of the likelihood under the constrained model, and $\ln L_1$ is the natural logarithm of the likelihood under the more complex, unrestricted model. The distribution of $\delta$ was assumed to approximate a $\chi^2$ distribution with $n$ degrees of freedom, where $n$ is the difference in the number of free parameters between the two nested models. The model containing the fewest parameters that did not differ significantly from the most general model was assumed to be the most appropriate description of the base-substitution process among the models examined. Each restricted model was tested against the most general model possible. The relationships among the six variants of the GTR model are shown in Figure 1.

Two points must be made regarding the use of the $\chi^2$ distribution in comparing nested among-site rate variation models. First, it is not possible to test the significance of $\delta$ between nonnested models by using a $\chi^2$ distribution (e.g., $GTR + SSR_{10}$ versus $GTR + I + \Gamma$). Second, as Whelan and Goldman (1999) have shown, the $\chi^2$ distribution may not be appropriate when testing the addition of among-site rate variation parameters. An alternative to both of these problems would be to use the parametric bootstrap to generate the null distribution of $\delta$ (Goldman, 1993). However, many of the models we examined are not included in available simulation software, and so we are unable to implement this approach here.

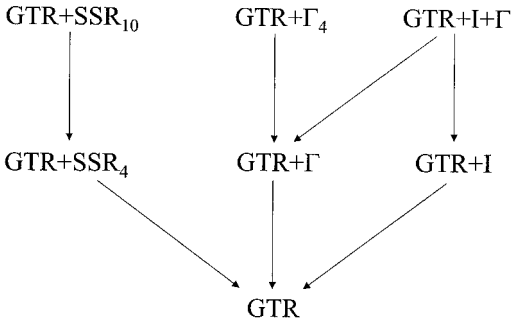



FIGURE 1. Hierarchical relationships among the six different among-site rate variation models used in this study. Models within the lower row are nested within those in the higher rows to which they are connected by lines. See text for descriptions of the assumptions made by the different models.

To investigate the effects of the high substitution rate at third positions on the relative fit of the various among-site rate variation models, we excluded the third positions and calculated the likelihood of the MP tree by using the three remaining character partitions. The four models evaluated were GTR, $GTR + I$, $GTR + \Gamma$, $GTR + I + \Gamma$, and an SSR model with three rate categories: one for the first positions, one for all second positions, and one for the $tRNA^{Asp}$ positions (referred to as the $GTR + SSR_3$ model).

Using MP, we calculated the percentage of varied sites within each of the codon positions and the $tRNA^{Asp}$ gene that had experienced more than one inferred substitution on the ML $GTR + I + \Gamma$ tree. Second, by excluding the constant sites from the data, we estimated SSRs on varied sites only for each of the above four character partitions (i.e., a modification of the $SSR_4$ model) to control for the different proportions of unvaried sites. We used the first approach as an estimate of the comparative distribution of rates among sites at those sites, which were observed to

vary among character partitions. The second approach was used as an estimate of the relative substitution rate at varied sites among character partitions.

### *Selection of an Optimal Tree Topology and Estimating Nodal Support*

Substitution model parameters were fixed in further ML tree searches. In all ML tree searches, a full heuristic search was undertaken with a starting tree obtained by stepwise addition (addition sequence simple), followed by TBR branch swapping. Starting branch lengths obtained by using the method of Rogers and Swofford (1998) were optimized ML under specific models of evolution. Trees were rejected if the approximate likelihood exceeded by >5% the best likelihood score encountered in the search. We did not search for the most likely tree under the GTR + $\Gamma_4$ model because PAML (Yang, 1997) does not implement the sophisticated tree-searching algorithms and is therefore prohibitively slow for the number of taxa we examined.

The ME trees were constructed under the same fully specified substitution models as the ML searches (except for the GTR + $\Gamma_4$ model). The distance estimators we used, (Waddell and Steel, 1997: their Equation 4) are referred to as ML distances by some authors (e.g., Swofford et al., 1996). Here, the R-matrix is optimized from all of the data and fixed for each pairwise comparison, unlike other, more commonly used distance estimators, where R is implicitly estimated in the calculation of each pairwise distance. Waddell and Steel (1997) found through simulation studies that homogenizing R for all distance comparisons tends to yield estimates with a lower variance. The disadvantage of this approach is its potential for bias if the true pattern of nucleotide substitution varies across the phylogeny. However, this assumption is also made by all commonly implemented ML methods (Felsenstein, 1981). These distance estimators are implemented in PAUP*4.0 by first defining a fully specified ML model and selecting "maximum likelihood distances" in the distance menu.

Nodal support was estimated by the nonparametric bootstrap method (Felsenstein, 1985) with 100 pseudoreplicates for the ML analyses and 500 pseudoreplicates for the ME analyses. The search strategy used was the same as described above for estimating topology. We performed ML bootstrapping under the GTR, GTR + I + $\Gamma$, and GTR + $SSR_{10}$ models only because of the computational burden of likelihood calculations. Bootstrapping was performed under all among-site rate variation models for the ME analyses except the GTR + $\Gamma_4$ model (because this model is restricted to PAML, which does not implement distance methods).

Because PAUP*4.0 (Swofford, 1998) ignores the codon position of a site during the generation of pseudoreplicates, we used the program CodonBootstrap 1.1 (J. Bollback, pers. comm.) to generate two data sets, one of 500 replicates and the other of 100 replicates, from the three protein-coding genes. CodonBootstrap 1.1 resamples codons instead of individual nucleotide sites, thus preserving the coding structure of the sequence. We pooled like codon positions from each of the three protein-coding genes and excluded the $tRNA^{Asp}$ data because CodonBootstrap 1.1 supports only three site-specific rate classes. Two bootstrap majority-rule consensus trees were then constructed from ME analysis of the 500-replicate data set and from ML analysis of the 100-replicate data set. Substitution model parameters were estimated from a uniform weighted MP topology and held constant in the analysis of each psuedoreplicate.

## RESULTS

### *Patterns of Nucleotide and Protein Evolution*

The alignment of sequences produced 1,520 homologous sites from each of the 25 individuals included in the phylogenetic analyses (available from the *Systematic Biology* website at: http://www.utexas.edu/ftp/depts/systbiol/). This alignment consisted of 753 bp from the 3′ end of the COI gene, the complete $tRNA^{Asp}$ (64 bp) and A8 (156 bp) genes, and 547 bp from the 5′ end of the A6 gene. For each species or subspecies, the individuals sampled were identical or nearly identical except for *M. campbelli* and *M. mangu*, in which we discovered putative cryptic species ("Otago" and "Awakino") (Buckley et al., 2001), and *M. cassiope* and *M. tenuis*, which also display geographic variation. These sequences have been deposited in GenBank under the following accession

numbers: AF247609 to AF247633, AF248797 to AF248820, AF248843, and AF249888. The A8 and A6 genes overlap by 7 bp, as they do in other insect mitochondrial genomes (e.g., Flook et al., 1995). We included these overlapping sites in the A8 character partition; however, all are unvaried in the sequences presented here. The secondary structure of the *M. campbelli* $tRNA^{Asp}$ gene is presented in Figure 2 to show the distribution of varied sites in the molecule.

We begin by presenting the inferred patterns of sequence variation within this data set, emphasizing the apparent complexity of the base-substitution process and how these processes appear to vary among character partitions. The sequence data presented here show patterns of variation typical of insect mtDNA (Simon et al., 1994), namely, transition bias, unequal distribution of varied/parsimony-informative sites among character partitions, and A + T richness (Table 1). Among the protein-coding gene regions, the A8 gene has the highest percentage of varied amino acid residues, followed by the A6 gene and finally the COI gene (Table 1), consistent with findings for other studies of insect mtDNA (Simon et al., 1994). The most extreme A + T richness is displayed by the A8 gene (87%), followed by $tRNA^{Asp}$ (80%), A6 (78%), and COI (73%) genes. However, when unvaried sites are excluded, the order of A + T richness becomes $tRNA^{Asp}$ (90%), COI (85%), A6 (84%), and A8 (82%). Thus, removing the confounding effects of unvaried sites reveals that the base composition at varied sites is homogenous among the protein-coding genes sampled here. Base composition varies greatly among the three codon positions (Table 1). Using the $\chi^2$ tests, we are unable to reject the hypothesis of base frequency stationarity among taxa for all sites, MP sites, or at any of the codon positions, or for any of the gene coding regions (Table 2). This observation is important because if the evolutionary process were nonstationary, then the ML estimation of substitution model parameters would be biased. However, the $\chi^2$ test for base composition stationarity is not very powerful because it ignores the correlation in base frequencies expected as a result of the phylogenetic relationships among taxa, and thus overestimates the number of degrees of freedom, increasing the chance of a type II error (however, an ME analysis using the logDet correction for nonstationarity produced an almost identical topology to many of the other stationary models). The base frequency parameters $\pi_A$, $\pi_C$, $\pi_G$, and $\pi_T$, obtained empirically, were 0.33, 0.12, 0.12, and 0.43, respectively, for all sites combined. These values were used in later likelihood calculations involving all of the sites.

The estimated percentage of varied sites that have experienced more than one substitution was inferred by using MP for first, second, and third positions for the three protein-coding genes combined and for the $tRNA^{Asp}$ gene. These values were used as crude measures of the overall substitution rate of each of the above four character partitions (Fig. 3). The varied third positions contain the greatest number of inferred multiple substitutions (57%), followed by the varied first positions (48%), $tRNA^{Asp}$ (31%), and finally the varied second positions (27%).

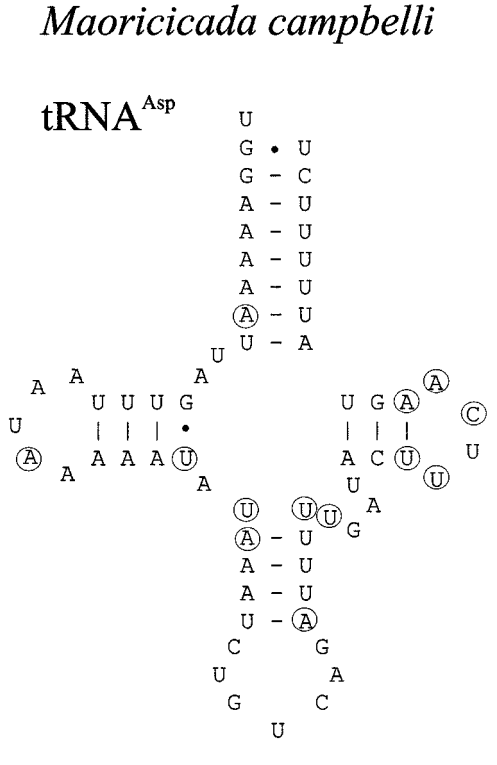



FIGURE 2. Predicted secondary structure of the *Maoricicada campbelli* $tRNA^{Asp}$ gene based on the structure from Syzmura et al. (1996), with varied bases circled.

TABLE 1. Sequence statistics and substitution model parameters for the various genes and character partitions. TS:TV ratios and $\alpha$ shape parameters are estimated under the HKY + $\Gamma$ model on the MP topology by using PAUP*4.0 (Swofford, 1998). Estimates of $\alpha$ were very similar under the GTR + $\Gamma$ model. Relative substitution rates are estimated under the GTR + $SSR_{10}$ model on the MP topology. nt = nucleotide, AA = amino acid, 2-fold = twofold degenerate, 4-fold = fourfold degenerate, TS:TV = transition/transversion ratio, SSRs = site-specific rates. Estimates are calculated from all taxa, including the outgroup species.

| | Character partitions | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | COI | | | | A6 | | | | A8 | | | | | |
| | 1st | 2nd | 3rd | All | 1st | 2nd | 3rd | All | 1st | 2nd | 3rd | All | tRNA | All sites |
| No. nt sites | 251 | 251 | 251 | 753 | 182 | 183 | 182 | 547 | 52 | 52 | 52 | 156 | 64 | 1520 |
| No. Varied nt sites | 26 | 3 | 165 | 194 | 47 | 12 | 114 | 173 | 14 | 14 | 27 | 55 | 13 | 435 |
| No. Parsimony nt sites | 14 | 1 | 107 | 122 | 19 | 6 | 67 | 92 | 9 | 6 | 19 | 34 | 5 | 253 |
| % nt sites varied | 10 | 1 | 66 | 26 | 26 | 7 | 63 | 32 | 27 | 27 | 52 | 35 | 20 | 29 |
| % nt sites MP | 6 | 0.4 | 43 | 16 | 10 | 3 | 37 | 17 | 17 | 12 | 37 | 22 | 8 | 17 |
| No. codons | | | | 251 | | | | 184 | | | | 52 | | 487 |
| Varied AA sites | | | | 19 | | | | 33 | | | | 23 | | 75 |
| Parsimony AA sites | | | | 6 | | | | 13 | | | | 13 | | 32 |
| % AA sites varied | | | | 7.6 | | | | 17.9 | | | | 44.2 | | 15.4 |
| % AA sites MP | | | | 2.4 | | | | 7.1 | | | | 25.0 | | 6.6 |
| % sites 2-fold | 7.2 | | 50.1 | 19.3 | 14.8 | | 47.2 | 20.7 | 1.9 | | 61.5 | 23.7 | | |
| % sites 4-fold | 0 | | 40.2 | 13.4 | 0 | | 31.9 | 10.6 | 0 | | 7.7 | 2.6 | | |
| Base freq. All sites | | | | | | | | | | | | | | |
| $\pi_A$ | 0.33 | 0.18 | 0.44 | 0.32 | 0.34 | 0.20 | 0.44 | 0.33 | 0.49 | 0.21 | 0.45 | 0.38 | 0.37 | 0.33 |
| $\pi_T$ | 0.32 | 0.44 | 0.46 | 0.41 | 0.40 | 0.47 | 0.49 | 0.45 | 0.41 | 0.57 | 0.48 | 0.49 | 0.43 | 0.43 |
| $\pi_C$ | 0.11 | 0.22 | 0.04 | 0.12 | 0.14 | 0.19 | 0.03 | 0.12 | 0.07 | 0.12 | 0.03 | 0.08 | 0.08 | 0.12 |
| $\pi_G$ | 0.24 | 0.16 | 0.06 | 0.15 | 0.12 | 0.13 | 0.04 | 0.10 | 0.03 | 0.09 | 0.04 | 0.05 | 0.11 | 0.12 |
| Base freq. Var. sites | | | | | | | | | | | | | | |
| $\pi_A$ | 0.42 | 0.04 | 0.42 | 0.42 | 0.33 | 0.39 | 0.42 | 0.39 | 0.60 | 0.09 | 0.27 | 0.31 | 0.46 | 0.39 |
| $\pi_T$ | 0.43 | 0.64 | 0.43 | 0.43 | 0.47 | 0.20 | 0.47 | 0.45 | 0.23 | 0.62 | 0.59 | 0.51 | 0.44 | 0.45 |
| $\pi_C$ | 0.14 | 0.01 | 0.07 | 0.08 | 0.12 | 0.33 | 0.05 | 0.09 | 0.05 | 0.16 | 0.07 | 0.09 | 0.08 | 0.08 |
| $\pi_G$ | 0.02 | 0.31 | 0.09 | 0.08 | 0.08 | 0.08 | 0.06 | 0.07 | 0.12 | 0.13 | 0.08 | 0.10 | 0.03 | 0.08 |
| TS/TV ratio | 20.17 | 3.25 | 7.27 | 6.68 | 4.62 | 1.73 | 13.35 | 8.09 | 14.03 | 6.03 | 7.03 | 7.35 | 3.17 | 6.97 |
| $\alpha$ shape parameter | 0.049 | 0.003 | 1.739 | 0.144 | 0.187 | 0.081 | $\infty$ | 0.186 | 0.151 | 0.398 | 1.091 | 0.295 | 0.126 | 0.168 |
| SSRs (all sites) | 0.235 | 0.014 | 2.734 | | 0.547 | 0.083 | 2.529 | | 0.976 | 0.463 | 1.878 | | 0.372 | |
| SSRs (varied sites) | 0.912 | 0.242 | 1.134 | | 0.597 | 0.210 | 1.227 | | 1.110 | 0.322 | 0.929 | | 0.602 | |

TABLE 2. Results from the $\chi^2$ tests for homogeneity of base composition among sequences. The $\chi^2$ tests are performed on all sites and on maximum parsimony-informative sites (MP) only.

| | *P*-values from $\chi^2$ base frequency test | |
|---|---|---|
| | All sites | MP sites only |
| All partitions | 1.000 | 0.628 |
| 1st positions | 0.858 | 0.734 |
| 2nd positions | 1.000 | 1.000 |
| 3rd positions | 1.000 | 0.802 |
| $tRNA^{Asp}$ | 1.000 | 0.999 |
| COI | 1.000 | 0.499 |
| A6 | 1.000 | 0.999 |
| A8 | 1.000 | 0.999 |

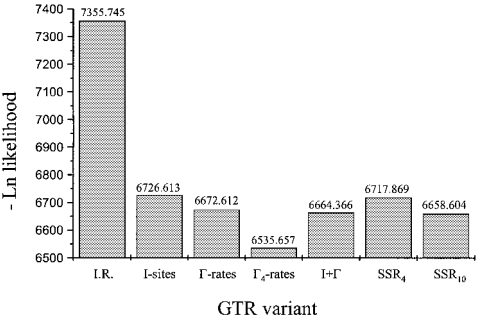



FIGURE 4. Relative fit of seven variants of the GTR model (GTR [I.R.], GTR + I, GTR + Γ, GTR + $\Gamma_4$, GTR + I + Γ, GTR + $SSR_4$, and GTR + $SSR_{10}$) to the MP topology (not shown).

### *Relative Fit of Substitution Models*

We used LRTs to critically evaluate different substitution models for subsequent phylogenetic reconstruction. Using this approach, we were able to reject all variants of the JC69, K80, and HKY85 models when tested for goodness of fit to the data against the variants of the GTR model. Of the GTR variants, the rank of best-fitting model, in order of decreasing negative log likelihood, is GTR (7355.745), GTR + I (6726.613), GTR + $SSR_4$ (6717.869), GTR + Γ (6672.612), GTR + I + Γ (6664.366), GTR + $SSR_{10}$ (6658.604), and GTR + $\Gamma_4$ (6535.657). The differences between the GTR + $SSR_{10}$ and GTR + $SSR_4$ models are significant ($\chi^2_{[6]} = 118.530$, $P < 0.05$), as are the differences between the GTR+I+Γ model and the GTR + I ($\chi^2_{[1]} = 124.49$, $P < 0.05$) and GTR+Γ ($\chi^2_{[1]} = 16.49$, $P < 0.05$) models. The invariable sites model has the poorest fit to the data of all the among-site rate variation models (Fig. 4).

The likelihood of the MP topology under the GTR + $\Gamma_4$ model was 6535.657, compared with 6672.612 for the GTR + Γ model. Using the LRT and $\chi^2$ tests, we can reject the GTR + Γ model in favor of the GTR + $\Gamma_4$ model at the 5% level of significance ($\chi^2_{[3]} = 273.91$, $P < 0.05$). Thus, we can reject the assumption that the distribution of among-site rate variation is homogeneous among the different codon positions and the $tRNA^{Asp}$ partition.

When third positions were excluded, the relative fit of the various GTR models to the data changed. The likelihood of the MP topology under the GTR + I, GTR + Γ, GTR + I + Γ, and GTR + $SSR_3$ models with third positions excluded were 2683.335, 2675.954, 2666.722, and 2793.615, respectively. Thus, of the above four among-site rate variation models, the $SSR_3$ model has the poorest fit to the data and the GTR + I + Γ has the best.

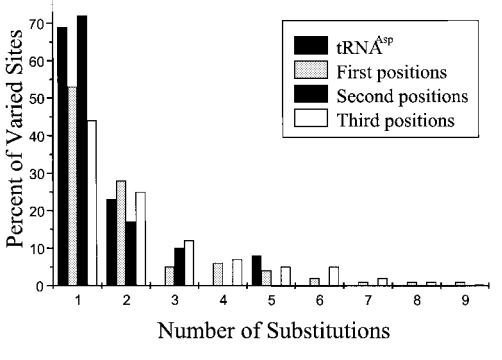



FIGURE 3. Inferred number of substitutions at each site for the three codon positions and the $tRNA^{Asp}$ gene estimated on the GTR + I + Γ ML tree (Fig. 6a). Percentages are calculated on varied sites only in order to correct for different proportions of unvaried sites within each of the four character partitions.

### *Substitution Model Parameter Estimates*

The estimation of substitution model parameters is a critical stage in model-based phylogenetic analysis. We have investigated how these estimates vary among character partitions and substitution model. The ML estimates of $\alpha$ shape parameters and TS:TV ratios reveals that the base-substitution process is heterogeneous among the various character partitions (Table 1). The $\alpha$ shape parameters estimated from the third positions are either >1 (COI and A8), or $\infty$ (A6),

indicating little or no among-site rate variation at third positions. In contrast, among-site rate variation at the first codon position appears to be extreme, as indicated by $\alpha$ shape parameters that are all $\leq 0.187$ (Table 1). The $\alpha$ values for the second codon positions of the COI and A6 genes are 0.003 and 0.081, respectively, also indicating even more extreme among-site rate variation. The $\alpha$ shape parameter for the A8 second positions is 0.398, the largest value of all the first and second position estimates. An $\alpha$ shape parameter of 0.126 for the $tRNA^{Asp}$ gene indicates considerable among-site rate variation in this gene also. The overall $\alpha$ value used for all sites in further likelihood tree searches under the GTR + $\Gamma$ model was 0.168 (Table 1). The value of $p_{inv}$ estimated under the GTR + I model was 0.683. The values of $p_{inv}$ and the $\alpha$ value estimated under the GTR + I + $\Gamma$ model were 0.555 and 0.711, respectively.

SSRs estimated under the GTR + $SSR_{10}$ model are presented in Table 1 and show that the average substitution rate is greatest at the third position, intermediate at the first position, and least at the second position for each of the three protein-coding genes (Table 1). When data for like sequence positions are pooled, the SSRs for the first, second, and third positions and the $tRNA^{Asp}$ sites are 0.416, 0.080, 2.588, and 0.377, respectively, under the GTR + $SSR_4$ model. The $tRNA^{Asp}$ sites have an average substitution rate, similar to that of the pooled first position sites. When SSRs are estimated on varied sites only, the pattern is the same, except that the magnitude of the difference in relative substitution rate (e.g.) between third and first positions, is not as great (Table 1). This observation indicates that unequal proportions of unvaried sites may inflate differences among the SSR parameters.

Under all the variants of the GTR model, a relative rate parameter describes the rate of each of the six possible nucleotide transformation types. The estimated values of these six parameters, $r_{CT}$, $r_{AG}$, $r_{AT}$, $r_{AC}$, $r_{CG}$, and $r_{GT}$ are given in Table 3. In this example, the $r_{GT}$ transversion is generally the rarest change; we set its rate at 1.0 and scaled the values of the other parameters relative to it. These estimates are highly dependent on the assumptions of the substitution model. The GTR, GTR + I, GTR + $\Gamma$, GTR + $\Gamma_4$, and GTR + I + $\Gamma$ models all infer that G $\leftrightarrow$ A transitions ($r_{AG}$) are the most frequent substitution type. On the other hand, the GTR + $SSR_4$ and GTR + $SSR_{10}$ models both infer an excess of C $\leftrightarrow$ T ($r_{CT}$) transitions over G $\leftrightarrow$ A transitions ($r_{AG}$). These two SSR models give very similar estimates for the relative rate parameters, and interestingly, those estimates are close to the values produced from the GTR model, which assumes equal rates among sites. The GTR + I model leads to the inference of higher substitution rates than does the GTR model for all substitution types except for T $\leftrightarrow$ A transversions ($r_{AT}$). The GTR + $\Gamma_4$ model gives the greatest substitution rates, with the more common transitions ($r_{CT}$ and $r_{AG}$) estimated to occur approximately three times more frequently than in the GTR model. Both the GTR + $\Gamma$ and GTR + $\Gamma_4$ models give similar estimates of the relative rate parameters.

### *Effect of Substitution Model Assumptions on Estimates of Topology, Branch Lengths, and Bootstrap Support Values*

Finally, we examined the effects of substitution model assumptions on estimates of bootstraps, topology, and branch lengths. Figure 5 shows branch lengths, estimated using ML, under the seven variants of the GTR

TABLE 3. Relative rate parameters from the R-matrices estimated under variants of the GTR model on the MP topology. The values given below were used in all further ML calculations under the respective substitution models.

| | Transitions | | Transversions | | | |
|---|---|---|---|---|---|---|
| | $r_{CT}$ | $r_{AG}$ | $r_{AT}$ | $r_{AC}$ | $r_{CG}$ | $r_{GT}$ |
| GTR | 30.550 | 35.559 | 3.924 | 2.845 | 0.965 | 1.0 |
| GTR + I | 38.121 | 43.105 | 2.590 | 3.233 | 1.691 | 1.0 |
| GTR + $\Gamma$ | 97.849 | 114.486 | 6.695 | 7.963 | 3.300 | 1.0 |
| GTR + $\Gamma_4$ | 98.611 | 107.181 | 6.245 | 7.567 | 3.477 | 1.0 |
| GTR + I + $\Gamma$ | 90.576 | 103.620 | 5.902 | 7.287 | 3.255 | 1.0 |
| GTR + $SSR_4$ | 34.734 | 31.550 | 2.328 | 2.727 | 1.844 | 1.0 |
| GTR + $SSR_{10}$ | 34.408 | 32.063 | 2.281 | 2.701 | 1.980 | 1.0 |

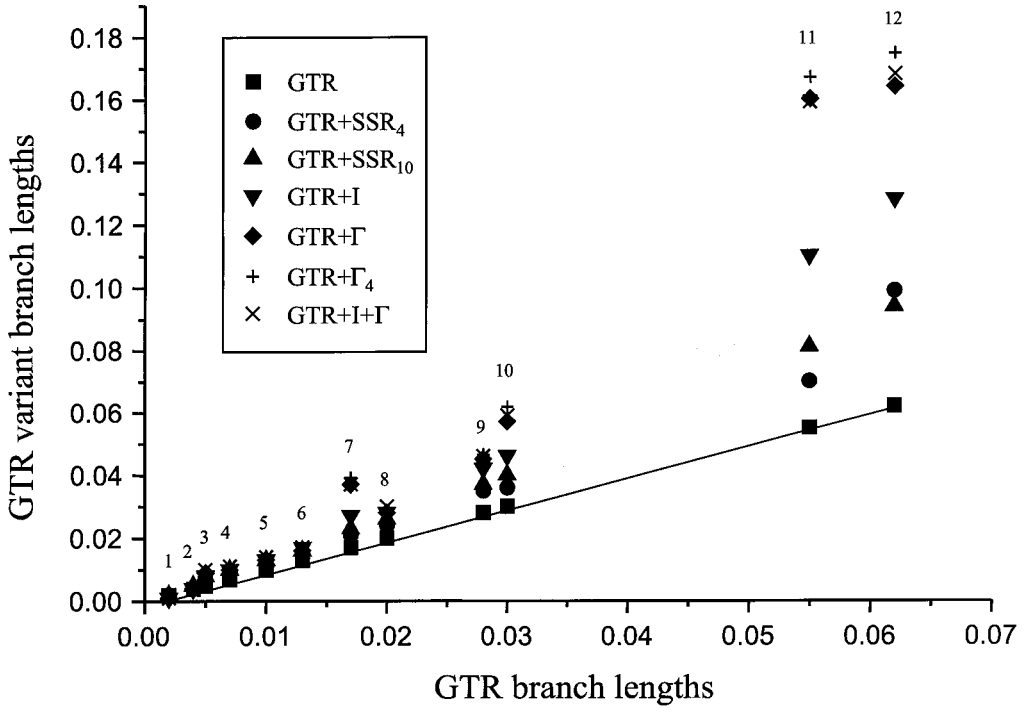
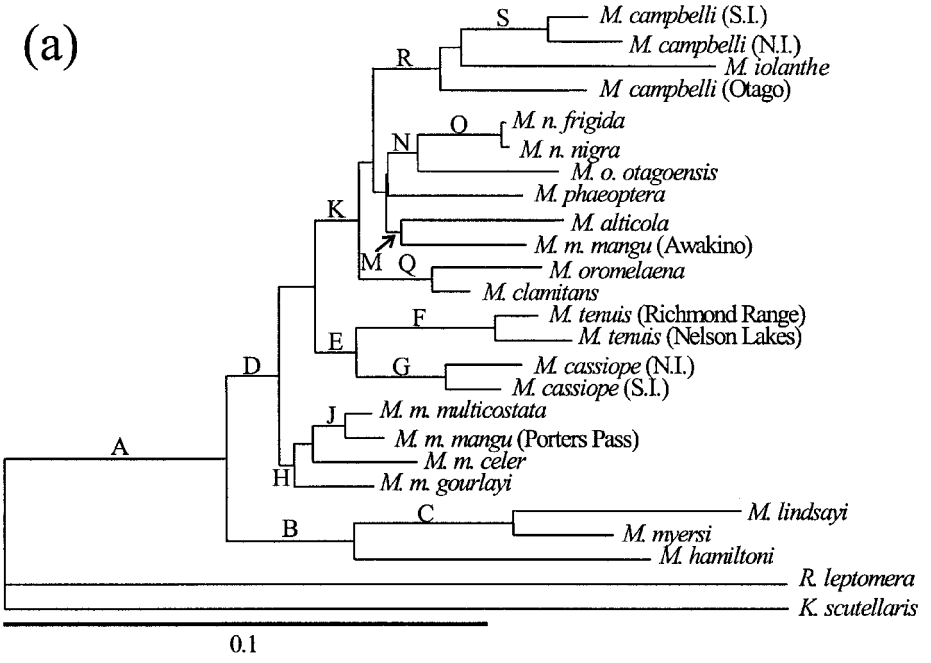



FIGURE 5. ML estimates of branch lengths under a range of variants of the GTR substitution model optimized on the MP topology. Each branch is labeled from 1 to 12 and is indicated on the tree in Figure 6b. The diagonal line connecting the GTR points on the *x* and *y*-axes illustrates the deviation of the among-site rate variation–corrected branch lengths from the equal rates estimates.

model, optimized on the MP topology. At low sequence divergence, each of the models gives very similar or identical branch lengths. The effect of accounting for among-site rate variation has a nonlinear effect on branch length estimates, with the longer branches increasing the most in inferred length relative to estimates by the equal rates model. For example, the longest branch length estimated under the GTR model is 0.062 substitutions per site, whereas the same branch has a length of 0.174 as estimated under the GTR + $\Gamma_4$ model. For the longer branches, the GTR model gives the lowest estimates, as expected. The two SSR models (GTR + $SSR_{10}$ and GTR + $SSR_4$) infer branch lengths that are similar to each other and closer to the GTR estimates than to those of the GTR + I, GTR + $\Gamma$, GTR + $\Gamma_4$, and GTR + I + $\Gamma$ models. The GTR + $\Gamma$, GTR + $\Gamma_4$, and GTR + I + $\Gamma$ models give the highest estimates for the long branches in this data set.

Despite the observation that branch lengths varied drastically between the different substitution models, the results of the phylogenetic analysis demonstrated a remarkable homogeneity in estimates of topology (Fig. 6). The ML trees varied in the placement of the *M. oromelaena* + *M. clamitans* clade and the placement of *M. phaeoptera*. Shifts of these lineages involved collapsing short internal branches and moving these branches to a neighboring node. None of the ML trees was identical to any of the ME trees. The ME trees differed from one another in their relative placement of *M. iolanthe*, *M. mangu*, and the *M. cassiope* + *M. tenuis* clade. As in the ML analyses, these topological changes involved only short internal branches and highly localized rearrangements. One notable difference among the substitution models was that the ML (Fig. 7) and ME invariable sites (data not shown) models placed the *M. hamiltoni*, *M. lindsayi*, and *M. myersi* clade as sister group to *M. cassiope* (ME tree) or to *M. cassiope* and *M. tenuis* (ML tree). This biologically unlikely placement (Buckley et al., 2001) renders *M. mangu* as paraphyletic and *M. m. gourlayi* as the basal *Maoricicada* species.



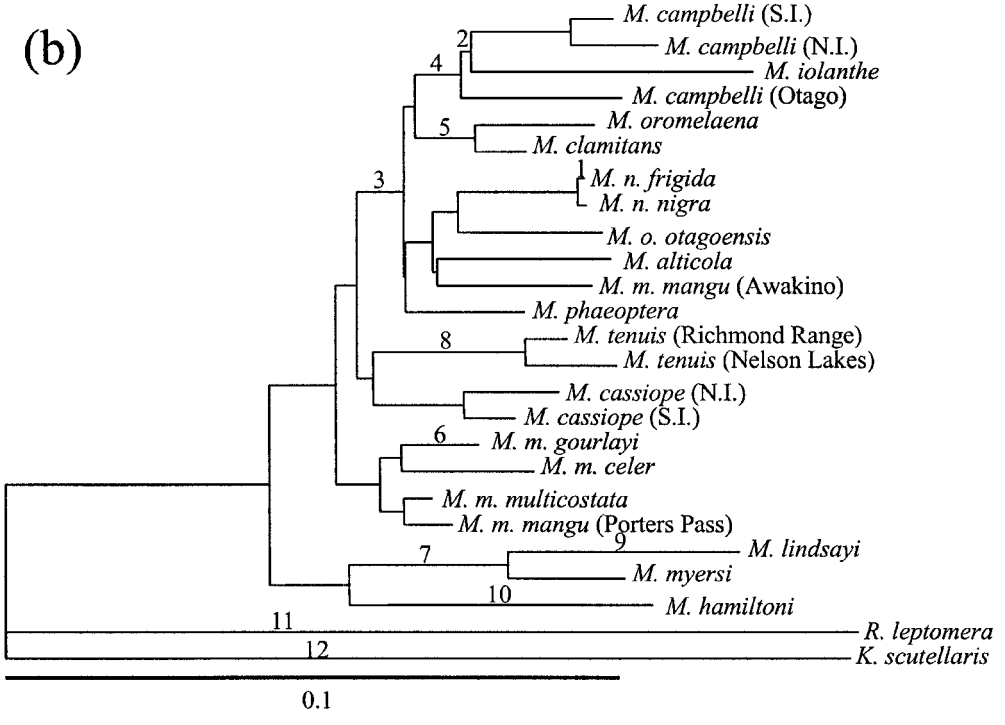




FIGURE 6. (a) Maximum likelihood and (b) minimum evolution trees estimated under the GTR + I + Γ model with branch lengths drawn proportionately to the expected number of substitutions per site. Letters above branches in (a) refer to the bootstrap proportions given in Tables 4 and 5. Numbers above branches in (b) refer to branches with lengths optimized in Figure 5. Note that branch 10 connects to the base of the *Maoricicada* radiation in the MP topology used to optimize branch lengths.

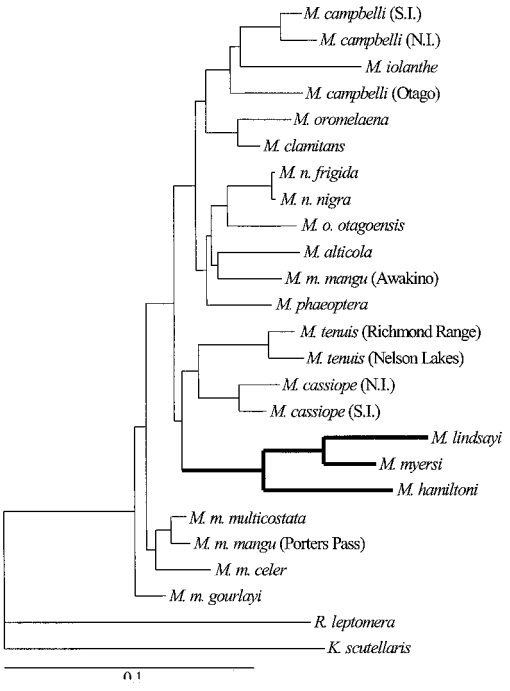



FIGURE 7. Maximum likelihood tree estimated under the GTR + I model, showing the placement of the *M. hamiltoni, M. lindsayi*, and *M. myersi* clade (indicated by thick branch lines) as sister group to *M. cassiope* and *M. tenuis*: a possible example of inaccuracy associated with an inappropriate model and adjacent long branches. The ME tree estimated under the GTR + I model gave a very similar topology, differing only in the placement of the two *M. tenuis* sequences (not shown). Branch lengths shown are proportional to the expected number of substitutions per site.

In addition to selecting the optimal topology, the estimation of bootstrap support is of critical interest. We observed that bootstrap support values for the entire data set varied appreciably among optimality criteria and substitution models for many nodes (Table 4). Variation in bootstrap values was particularly evident for nodes uniting long internal branches. For example, for the branch that partitions *M. hamiltoni*, *M. myersi*, and *M. lindsayi* and the two outgroup species from the rest of the taxa included in the tree (D in Fig. 6a), estimates of bootstrap support ranged from 98% (equal rates) to 27% (I-sites) for ME and from 83% (equal rates) to 47% (I + Γ) for ML (Table 4). Note that this same node was not recovered in the optimal ML (Fig. 7) or the ME GTR + I (data not shown) trees. Another node that received particularly variable bootstrap support was the one uniting the *M. mangu* subspecies (except for the *M. m. mangu* sequence from Awakino). Bootstrap support for this node ranged from 100% (equal rates, $SSR_4$ and $SSR_{10}$) to 52% (I-sites) in the ME analyses and from 78% (equal rates) to 30% (I + Γ) for ML (Table 4). Other nodes have more consistent estimates of bootstrap support from the various substitution models. For example, the node uniting *M. oromelaena* and *M. clamitans* was supported by bootstrap proportions ranging from 100% ($SSR_{10}$) to 97% (I + Γ) under ME and from 99% (I + Γ) to 97% (equal rates) under ML. Some nodes

TABLE 4. Minimum evolution (ME) and maximum likelihood (ML) bootstrap support values for various nodes estimated under variants of the GTR model. For the ME analyses, bootstrap values were estimated under equal rates (I.R.), a four rate-class site-specific rates model ($SSR_4$), 10 rate-class site-specific rates model ($SSR_{10}$), invariable sites (I-sites), gamma rates (Γ-rates), and a mixed invariable sites and gamma rates model (I + Γ). For the ML analyses, bootstrap values were estimated under the I.R., $SSR_{10}$, and I + Γ models only. Nodes are labeled from A to S and are as indicated on the tree in Figure 6a.

| | Minimum evolution | | | | | | Maximum likelihood | | |
|---|---|---|---|---|---|---|---|---|---|
| Node | I.R. | $SSR_4$ | $SSR_{10}$ | I-sites | Γ-rates | I + Γ | I.R. | $SSR_{10}$ | I + Γ |
| A | 99 | 100 | 99 | 83 | 100 | 96 | 100 | 99 | 85 |
| B | 81 | 82 | 82 | 85 | 83 | 87 | 79 | 80 | 86 |
| C | 100 | 100 | 100 | 98 | 100 | 99 | 100 | 100 | 100 |
| D | 98 | 96 | 96 | 27 | 64 | 39 | 83 | 57 | 47 |
| E | 47 | 43 | 43 | 30 | 45 | 44 | 52 | 54 | 65 |
| F | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 99 | 100 |
| G | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 99 | 99 |
| H | 100 | 100 | 100 | 52 | 81 | 65 | 78 | 56 | 30 |
| J | 93 | 96 | 95 | 74 | 70 | 72 | 87 | 85 | 85 |
| K | 69 | 74 | 72 | 43 | 56 | 46 | 61 | 61 | 65 |
| M | 35 | 43 | 35 | 38 | 39 | 40 | 35 | 42 | 54 |
| N | 68 | 77 | 70 | 72 | 72 | 73 | 66 | 65 | 59 |
| O | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Q | 99 | 99 | 100 | 98 | 98 | 97 | 97 | 98 | 99 |
| R | 86 | 86 | 86 | 89 | 91 | 91 | 82 | 92 | 91 |
| S | 100 | 100 | 100 | 100 | 99 | 100 | 100 | 100 | 100 |

were supported by >95% from both ML and ME under all substitution models (e.g., nodes C, F, G, O, Q, and S in Fig. 6a). These nodes connected pendant branches between closely related species (C and Q) or among populations within a species (G, F, O, and S).

The bootstrap replicates generated by using PAUP*4.0 (Swofford, 1998) yielded the same ME majority rule consensus topology as the replicates generated by using CodonBootstrap 1.1 (provided by J. Bollback). Estimates of nodal bootstrap support were also very similar between the two programs, with a maximum difference of 8% for nodes E and M (Table 5). Three nodes received greater support from PAUP*, four nodes received greater support from CodonBootstrap 1.1 and nine nodes had equal estimates from both programs. For the ML analyses the biggest difference in nodal support was 25% for node D (Table 5). This node also had highly labile estimates of bootstrap support for the full data set under a wide range of substitution models (Table 4). Seven nodes received greater support in the PAUP* analysis, six nodes received greater support in CodonBootstrap 1.1, and three nodes had equal support in both programs. Under both ML and ME, neither program gave bootstrap support values that were consistently biased towards or away from the equal rates bootstrap estimates.

## DISCUSSION

As is now well established, among-site rate variation can have extremely detrimental effects on several aspects of phylogenetic analysis (reviewed by Yang, 1996b). This problem can be addressed by using an appropriate substitution model during tree selection. However, the molecular systematist is faced with a wide variety of substitution models from which to choose. We have compared the performance of six different types of substitution models that explicitly account for among-site rate variation (I-sites, $\Gamma$-rates, $\Gamma_4$, $I+\Gamma$, $SSR_4$, and $SSR_{10}$ models) by using a typical molecular systematics data set (i.e., 25 taxa, 1,520 sites, and 253 parsimony-informative sites). Although each of these six among-site rate variation models and the equal rates model lead to selection of very similar optimal topologies for our particular data set, we observed large differences between the models in estimates of R-matrices, branch lengths, and bootstrap proportions.

We observed that both SSR models, $SSR_{10}$ and $SSR_4$, gave lower estimates of the rate of the more commonly occurring transition type substitutions relative to the I-sites, $\Gamma$-rates, and $I+\Gamma$ models (Table 3). The values inferred from the SSR models were almost as low as those from the equal rates model, which is known to be a highly biased estimator of the pattern of nucleotide

TABLE 5. Minimum evolution and maximum likelihood bootstrap support values for various nodes estimated under variants of the site-specific and equal rates (I.R.) models. Bootstrap replicates were generated by using PAUP*4.0 (Swofford, 1998) or CodonBootstrap 1.1 (J. Bollback, pers. comm.) and trees were estimated by using PAUP*4.0. Nodes are labeled as in Figure 6a. Five hundred replicates were analyzed under minimum evolution and 100 replicates under maximum likelihood.

| | Minimum evolution | | | Maximum likelihood | | |
|---|---|---|---|---|---|---|
| Node | I.R. | PAUP* | Codon bootstrap | I.R. | PAUP* | Codon bootstrap |
| A | 100 | 100 | 100 | 100 | 100 | 97 |
| B | 77 | 86 | 82 | 63 | 63 | 71 |
| C | 100 | 100 | 100 | 100 | 100 | 100 |
| D | 96 | 98 | 98 | 86 | 86 | 61 |
| E | 49 | 55 | 47 | 70 | 70 | 58 |
| F | 100 | 100 | 100 | 100 | 100 | 100 |
| G | 100 | 100 | 100 | 100 | 100 | 96 |
| H | 100 | 100 | 100 | 80 | 80 | 69 |
| J | 97 | 97 | 98 | 86 | 86 | 89 |
| K | 76 | 66 | 74 | 46 | 46 | 66 |
| M | 39 | 44 | 43 | 38 | 38 | 35 |
| N | 66 | 61 | 64 | 40 | 40 | 44 |
| O | 100 | 100 | 100 | 100 | 100 | 99 |
| Q | 100 | 99 | 100 | 97 | 97 | 99 |
| R | 87 | 88 | 88 | 81 | 81 | 82 |
| S | 100 | 100 | 100 | 100 | 100 | 100 |

substitution (Wakeley, 1994, 1996). We also observed that both of the SSR models yielded branch length estimates that were less than those from the I-sites, Γ-rates, and I + Γ models and closer in magnitude to the equal rates estimates. The relative underestimation of R-matrices and branch lengths by the SSR models relative to the I-sites, Γ-rates, and I + Γ models may result from the assumption made by the SSR model that each site within a given rate class is equally likely to accept a substitution, an assumption that may lead to underestimates of the number of multiple substitutions within rate classes where among-site rate variation is extreme (e.g., at first and second positions; Table 1). This hypothesis is supported by the observation that the likelihood scores of the I-sites, Γ-rates, and I + Γ models were improved relative to the $SSR_3$ model when third positions were removed from the analysis. Apparently the superior likelihood score of the $SSR_{10}$ model can be attributed to the inflated substitution rate at third positions. Thus, a gamma model or an invariable sites model (or a combination) will probably always be better for describing the distribution of rates among sites for first and second codon positions.

If the SSR models are inaccurate in correcting for multiple hits, then we expect to observe that the relative frequencies of transitions are underestimated in the R-matrices and their corresponding branch lengths. We believe that the invariable sites and gamma rates models are giving more accurate estimates of branch lengths because the SSR model estimates are very close to the equal rates values, which previous authors have shown to be highly biased (Yang, 1996b). Waddell and Steel (1997) observed that for a data set of hominoid mtDNA, the gamma model inferred greater TS:TV ratios and distances than did the invariable sites model, which is in agreement with our observations. We suspect that the gamma model will in general infer longer branch lengths than the invariable sites model, except in extreme examples (see Waddell et al., 1997). Because the invariable sites model assumes a constant rate for all variable sites, no sites have an extremely high rate under this model. Similarly, the SSR model will have a fixed upper substitution rate, which will be determined by the fastest rate category. Under a gamma model, however, there is a much higher upper limit for the rate at which a site can evolve; if we were not using discrete gamma models, this upper limit would be infinity (P. Lewis, pers. comm.).

The underestimation of the true number of substitutions that have occurred along a branch can manifest itself in "long branch attraction" (Felsenstein, 1978; Hendy and Penny, 1989) even when ML estimation is used, if the model of evolution fits the data poorly (e.g., Sullivan and Swofford, 1997). Although we noted some differences among models in terms of selection of an optimal topology, these differences all tended to be restricted to nodes that were poorly supported in the bootstrap analyses. Well-supported nodes tended to be recovered by all models and both optimality criteria. If the correct topology is to be obtained, the correct optimization of branch lengths is more important in lineges with high rates of change or extreme rate variation (Cunningham et al., 1998). Because neither of these two phenomena seems to characterize the data presented here, we are not surprised that the process of tree selection is relatively insensitive to the substitution model in this example. Such will not always be the case, for example, with higher rates of change (deeper divergences) or appreciable differences in rates of change between lineages (e.g., Sullivan and Swofford et al., 1997; Cunningham et al., 1998; Takezaki and Gojobori, 1999).

In addition to incorrect branch length estimation causing long branch attraction, overestimation of the number of multiple substitutions on trees where long branches are correctly united may throw ML (when used with an inappropriate model) into a region of tree space where long branches will separate from each other (Waddell et al., 1999). We have identified a possible example of this poorly understood phenomenon in the biologically unlikely (Buckley et al., 2001) placement of the *M. hamiltoni*, *M. myersi*, and *M. lindsayi* clade, in both the ML (Fig. 7) and ME trees estimated under the GTR + I model. In this situation the invariable sites model may have overadjusted for among-site rate variation, leading to misplacement of the long branch forming the *M. hamiltoni, M. myersi*, and *M. lindsayi* clade relative to the long outgroup branch. If so, then both likelihood and distance-based methods (see also Bruno et al., 2000) are apparently susceptible to this form of systematic error.

When sites are evolving under a distribution of rates, that is, where some sites are invariable and the remainder fall under a range of constraints (i.e., as approximated by an $I + \Gamma$ model), then $p_{inv}$ may be overestimated because the invariable sites model assumes that what are actually slowly evolving sites are instead invariable. Because the data presented here fit a model assuming a range of substitution rates (i.e., the $I + \Gamma$ assumption) better than they fit an invariable sites model, we believe that the above explanation is the best. The invariable sites model may be expected to perform better when the distribution of rates among sites is closer to a true bimodal form (Waddell et al., 1997). We also observed that using a lower value of $p_{inv}$ (i.e., $p_{inv}$ as estimated under the $I + \Gamma$ model) led to selection of a topology (data not shown) where the *M. hamiltoni*, *M. myersi*, and *M. lindsayi* clade is basal to the remaining *Maoricicada* species, which is in agreement with the other substitution models (Fig. 6) and exemplifies the importance of accurate parameter optimization. This observation is not unique to the invariable sites model. If we construct ME and ML trees under a GTR + $\Gamma$ model and using an $\alpha$ value lower than that estimated by ML, we get similar topologies (data not shown) to that obtained under the GTR + I model—again underscoring the importance of using an appropriate substitution model and accurately estimating the parameters of this model.

Further studies are required to determine whether the trends observed in estimates of R-matrices, branch lengths, and topology are general properties of the SSR model or are restricted to the data presented here. Simulation studies and tests on "known" phylogenies will be particularly informative in this respect. The most appropriate model for the accurate reconstruction of topology is likely to be data set–specific.

Despite the fact that most substitution models led to the selection of similar topologies, bootstrap support varied drastically for some nodes. This variation was particularly evident at nodes connecting long internal branches. As parameters are added to a substitution model, bootstrap support for a node may increase because of an increase in the accuracy of tree selection. Conversely, parameter addition may also cause bootstrap support to drop, for two reasons. First, overly simplistic substitution models may underestimate the true number of multiple substitutions that have occurred on long branches; thus giving inflated bootstrap support for such nodes (Sullivan et al., 1997). Second, parameter-rich substitution models typically have a higher variance than models with fewer parameters (Kumar et al., 1993), which is often reflected in decreasing estimates of nodal bootstrap support (Waddell and Steel, 1997). However, comparing the variances of different estimators is not meaningful when one of the estimators is biased or inconsistent, which many simple substitution models are (Yang, 1994c).

In our analyses, we noted that bootstrap values both increased and decreased when we accounted for among-site rate variation, as has been observed in other studies (Sullivan et al., 1997). For example, in the ME analyses, bootstrap support for five of the labeled nodes in Figure 6a increased under the GTR + I + $\Gamma$ model relative to the GTR model, whereas support for eight nodes decreased and that for five other nodes remained constant. Similarly, in the ML analyses, bootstrap support increased for six nodes, decreased for six other nodes, and remained constant for four other nodes. Reyes et al. (1998) argued that the use of a parameter-rich model might cause statistical fluctuations that will prevent any significant phylogeny being inferred from any given set of sequences. However, we have shown that the relationship between the number of parameters in a substitution model and estimates of bootstrap support is far from simple. We subscribe to the position (e.g., Corneli and Ward, 2000) that biological realism should not be sacrificed for the convenience of performing phylogenetic analyses under the assumptions of a simple substitution model. In addition, if the model is overly simplistic or biased, strong support for an incorrect phylogenetic hypothesis can be inferred (Lockhart et al., 1996; Sullivan and Swofford, 1997). The use of a more complex model, although perhaps incapable of giving strong support for the "true" hypothesis, may yet reveal that the difference among the competing hypotheses is nonsignificant. We believe that the latter situation is certainly preferable to the former and thus advocate the use of realistic substitution models in phylogenetic analysis wherever possible.

Because the GTR + $SSR_{10}$ model uses a large number of parameters to describe the

distribution of among-site rate variation, the estimation of parameter values is potentially more susceptible than a less complex model to random error. This can be partially compensated for by increasing the number of sequences in the analysis (e.g., Sullivan et al. 1999) but, of course, this is not always possible. However, as we noted in Table 4, for the ML GTR + $SSR_{10}$ and ML GTR + I + $\Gamma$ analyses, although estimates of bootstrap support differed for many nodes, neither method gave consistently higher or lower values. Thus, despite the fact that the GTR + $SSR_{10}$ model has 17 free parameters compared with 10 for the GTR + I + $\Gamma$ model, variance in the estimation of model parameters does not seem to be reducing the signal in the data to nonsignificant values. The effects of increasing the number of parameters in the estimation of branch lengths, topology, and bootstrap proportions have not previously been well studied (but see Corneli and Ward, 2000). The increasing use of longer DNA sequences (>1 kb) renders this issue less important than that of obtaining an improved fit between the model assumed and the true underlying evolutionary process (Swofford et al., 1996; Sullivan and Swofford, 1997). Of equal interest is the relationship between the complexity of the selected model and the associated computational burden of likelihood calculations. However, the continued development and refinement of algorithms (e.g., Lewis, 1998) and the parallelization of phylogenetic analysis software offer hope that parameter optimization and tree searches on large data sets and using complex models will become increasingly feasible.

We assessed the potential bias in estimating bootstrap support under an SSR model when the codon structure of a protein-coding gene is ignored during resampling. Our results indicate only slight differences between PAUP* and CodonBootstrap 1.1 for the ME analyses (Table 5). Correcting for among-site rate variation under an SSR model had little effect on estimates of bootstrap support for nodes in trees produced from this data set anyway, and thus we are not surprised by this observation. Under ML, some nodes showed large differences in estimates of bootstrap support, although the small number of replicates that we were able to perform, given computational constraints, may have exaggerated these differences. Neither method of generating bootstrap replicates seems to be biased in any particular direction with respect to the equal rates bootstrap estimates.

An alternative explanation for the variation in bootstrap support among different models is that a bias may have been introduced into the ML heuristic search strategy. We used single round of TBR branch swapping from a stepwise addition tree obtained via a simple addition sequence. Depending on the nature and complexity of tree space, such a search may be prone to entrapment in local optima. To examine this potential bias we repeated the ML bootstrap analysis under the GTR model. Each bootstrap replicate involved a starting tree obtained by using stepwise addition with random addition sequence and 10 replicates. Although we would have preferred to use this search strategy for all of our ML bootstrap analyses, the large computational burden precluded this. Fourteen of the 16 nodes marked on the phylogeny in Figure 6a differed by <5% between the two search strategies. Only two nodes (K and N) received appreciably different values, 49% and 59%, respectively. Thus, although the search strategy appears to have had some effect on the variation in bootstrap support among models, we do not believe it is responsible for the large differences observed among models for some nodes. The small number of replicates we were able to analyze may have also exaggerated the differences among models, although as the above analyses indicate, this effect is not large either. For the ME bootstraps we used 500 replicates, so the number of replicates is not an issue for those analyses.

The SSR model requires that all sites be partitioned into a prespecified number of rate classes. Within each rate class, all sites are assumed to share an identical substitution rate that is different from that of other such rate classes. However, the process of identifying and characterizing these rate classes can be problematic. For protein-coding genes, the division of sites according to codon position may seem an obvious way to delineate rate classes; however, the distribution of rates among sites may show an extensive overlap between codon positions, as we have shown in Figure 3 (see also Olmstead et al., 1998). Also among-site rate variation may be extreme within codon positions, especially the first and second positions (see Table 1; Yang, 1996c; Voelker and Edwards, 1998). Thus, although the SSR model is more realistic than

an equal rates model (e.g., the GTR model), the assumption of the SSR model that all sites within each rate class are evolving at the same rate is clearly violated in the example presented here—and probably always will be found to be. When analyzing protein-coding genes, substitution models based around the codon (e.g., Goldman and Yang, 1994; Muse and Gaut, 1994) may be preferable because they take into account the degeneracy of a site.

Although we have pooled all tRNA sites into a single rate category, this may be an inappropriate method of dealing with RNA coding sequences, although in our study it probably had little effect because the number of base pairs involved is very small. Genes coding for tRNA molecules are expected to contain both variable and conserved (e.g., the anticodon) sites. Partitioning tRNA and rRNA sites into rate categories according to their inclusion in a helix or unpaired region is likely to be inappropriate because conserved and highly variable sites are known to exist in both (e.g., Sullivan et al., 1995; Hickson et al., 1996; Fig. 2; Buckley et al., 2000). Additionally, noncoding regions such as introns, the mitochondrial control region, and nuclear internal-transcribed spacer sequences have poorly known functional constraints. With little knowledge regarding the nature of these real or hypothesized constraints, it is difficult to identify rate classes from observed patterns in the data, except when very large numbers of sequences are available (e.g., Van de Peer et al., 1993).

The gamma distribution is an attractive model for describing among-site rate variation for two major reasons (reviewed by Yang, 1996b). First, the distribution of rates among sites is described by a single parameter, $\alpha$. The $\alpha$ shape parameter is comparable across data sets, provided the same number of rate categories are used (Yang, 1994b), which allows generalizations to be made regarding the relative extent of among-site rate variation among genes. The relative substitution rates estimated under the SSR model, although comparable among character partition within the same data set, are not comparable across studies. Second, the gamma distribution is versatile in that it can accommodate a wide range of site-rate distributions, ranging from a distribution that is nearly homogeneous to one that is highly variable (i.e., containing a large proportion of sites evolving at a rate close to zero combined with a proportion of hypervariable sites). As has been shown in this and other studies (Gu et al., 1995; Waddell et al., 1997), adding a class of sites with a rate of zero (i.e., an $I + \Gamma$ model) can further extend and improve the gamma model.

Another approach to among-site rate variation is the Hidden Markov Model (HMM) of Felsenstein and Churchill (1996). We have not used this approach, because as it is currently implemented, the HMM cannot be used in conjunction with the GTR model; moreover, it has been used in very few empirical studies (Cook et al., 1999). However, this approach has some attractive features. First, the HMM is not restricted by the assumption that we should know the relative rate at a particular site. Second, the HMM is not restricted by the assumption that the distribution of rates among sites follows a gamma distribution, and it can also include a class of invariable sites. Third, the HMM can account for the spatial correlation in substitution rate among sites, with the addition of another parameter, although Yang's (1995) space–time process model, also a hidden Markov model, allows for the spatial correlation of rates. We believe this model presents a powerful alternative to the approaches implemented here and hope that the HMM will be more thoroughly explored in future empirical and theoretical studies.

A further potential problem to modeling among-site rate variation by using the approaches discussed here is the possibility of a change in the distribution of sites that are free to vary among sequences. Such covariotide shifts (Fitch and Markowitz, 1970; Waddell et al., 1997; Lockhart et al., 1998) may confound any model of evolution that assumes any given site has the same probability of change over all sequences. Although shifts in the distribution of among-site rate variation at the level of the codon (covarion shift) and individual nucleotide site (covariotide shift) have been shown to occur in deep-level phylogenies (e.g., Miyamoto and Fitch, 1995; Lockhart et al., 1998), the importance of such shifts for recently diverged taxa, such as the example presented here, is probably much less but to date has been largely unexplored. Shifts in base frequency parameters (Galtier and Gouy, 1998) and TS:TV ratios (Yang and Yodder, 1999) can now be accommodated in ML; however, work on modeling covariotide

processes is in its infancy (Tuffley and Steel, 1998).

We advocate testing phylogenetic hypotheses under a range of substitution models as a method of data exploration. No current model accounts for all of the vagaries of nucleotide substitution; all are biased to some extent. By exploring the effects of these assumptions on the analysis, the molecular systematist can gain a greater understanding of both the data they obtain and the models used to analyze these data.

## ACKNOWLEDGMENTS

We are very grateful to all the people who provided specimens, information regarding collecting localities, assistance in fieldwork, and laboratory work: Peter Arensburger, Wee Ming Boon, Bert Borger, Steve Chiswell, Darren Day, Kathy Donohue, John Dugdale, Peg Fleming, Francesco Frati, George Gibbs, Dianne Gleeson, David Lane, Mathew Lark, Liz MacAvoy, Mary McEwen, Lesley Milicich, and Paul Sunnucks. This work benefited greatly from discussions or communications with Jonathan Bollback, John Huelsenbeck, Peter Lockhart, David Penny, Andrew Rambaut, Ted Schultz, Jack Sullivan, Mike Steel, David Swofford, Simon Whelan, and Ziheng Yang. Thanks also to Jonathan Bollback for providing the CodonBootstrap program. Paul Lewis, Keith Crandall, Richard Olmstead, and two anonymous reviewers gave comments on the manuscript. We appreciate the efforts of the New Zealand Department of Conservation (Te Papa Atawhai) for providing collecting permits and advice on access to field sites. Throughout this study, we have been supported by grants from the National Geographic Society (to C.S., G.K.C., George Gibbs, and Maxwell Moulds), Victoria University of Wellington (T.R.B. and G.K.C.), the National Science Foundation (C.S.), the Fulbright Foundation (C.S.), and the University of Connecticut (C.S.).

## REFERENCES

BRUNO, W. J., N. D. SOCCI, AND A. L. HALPERN. 2000. Weighted neighbor joining: A likelihood-based approach to distance-based phylogeny reconstruction. Mol. Biol. Evol. 17:189–197.

BUCKLEY, T. R., C. SIMON, P. K. FLOOK, AND B. MISOF. 2000. Secondary structure and conserved motifs of the frequently sequenced domains IV and V of the insect mitochondrial large subunit rRNA gene. Insect Mol. Biol. 9:565–580.

BUCKLEY, T. R., C. SIMON, H. SHIMODAIRA, AND G. K. CHAMBERS. 2001. Evaluating hypotheses on the origin and evolution of the New Zealand alpine cicadas (*Maoricicada*) using multiple comparison tests of tree topology. Mol. Biol. Evol., in press.

CABOT, E. L., AND A. T. BECKENBACH. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput. Appl. Biosci. 5:233–234.

COOK, C. E., Y. WANG, AND G. SENSABAUGH. 1999. A mitochondrial control region and cytochrome *b* phylogeny of sika deer (*Cervus nippon*) and report of tandem repeats in the control region. Mol. Phylogenet. Evol. 12:47–56.

CORNELI, P. S., AND R. H. WARD. 2000. Mitochondrial genes and mammalian phylogenies: Increasing the reliability of branch length estimation. Mol. Biol. Evol. 17:224–234.

CUNNINGHAM, C. W., H. ZHU, AND D. M. HILLIS. 1998. Best-fit maximum likelihood models for phylogenetic inference: Empirical tests with known phylogenies. Evolution 52:978–987.

DUGDALE, J. S. 1972. Genera of New Zealand cicadidae (Homoptera). N.Z. J. Sci. 14:856–882.

DUGDALE, J. S., AND C. A. FLEMING. 1978. New Zealand cicadas of the genus *Maoricicada* (Homoptera: Tibicinidae). N.Z. J. Zool. 5:295–340.

EDWARDS, A. W. F. 1992. Likelihood, expanded edition. Cambridge Univ. Press, Cambridge.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

FELSENSTEIN, J. 1993. PHYLIP: Phylogeny inference package, version 3.5c. Dept. of Genetics, Univ. of Washington, Seattle.

FELSENSTEIN, J., AND G. A. CHURCHILL. 1996. A hidden Markov model approach to variation among sites in rate of evolution. Mol. Biol. Evol. 13:93–104.

FITCH, W. M. 1971. Toward defining the course of evolution: Minimum change for a specific tree topology. Syst. Zool. 20:406–416.

FITCH, W. M., AND E. MARKOWITZ. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem. Genet. 4:579–593.

FLEMING, C. A. 1971. A new species of cicada from rock fans in Southern Wellington, with a review of three species with similar songs and habitat. N.Z. J. Sci. 14:443–479.

FLOOK, P., C. H. F. ROWELL, AND G. GELLISSEN. 1995. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. 41:928–941.

FRATI, F., C. SIMON, J. SULLIVAN, AND D. L. SWOFFORD. 1997. Evolution of the mitochondrial cytochrome oxidase II gene in collembola. J. Mol. Evol. 44:145–158.

GALTIER, N., AND M. GUOY. 1998. Inferring pattern and process: Maximum likelihood implementation of a nonhomogenous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. 15:871–879.

GAUT, B. S., AND P. O. LEWIS. 1995. Success of maximum likelihood inference in the four-taxon case. Mol. Biol. Evol. 12:152–162.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36:182–198.

GOLDMAN, N., AND Z. YANG. 1994. A codon-based model of nucleotide substitution for protein coding DNA sequences. Mol. Biol. Evol. 11:725–736.

GU, X., Y.-X. FU, AND W.-H. LI. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol. Biol. Evol. 12:546–557.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating of the human–ape split by a molecular clock by mitochondrial DNA. J. Mol. Evol. 22:160–174.

HENDY, M. D., AND D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38:297–309.

HICKSON, R. E., C. SIMON, A. COOPER, G. S. SPICER, J. SULLIVAN, AND D. PENNY. 1996. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. Mol. Biol. Evol. 13:150–169.

HILLIS, D. M., J. P. HUELSENBECK, AND D. L. SWOFFORD. 1994. Hobgoblin of phylogenetics. Nature 369:363–364.

HUELSENBECK, J. P. 1995a. Performance of phylogenetic methods in simulation. Syst. Biol. 44:17–48.

HUELSENBECK, J. P. 1995b. The robustness of two phylogenetic methods: Four taxon simulations reveal a slight superiority of maximum likelihood over neighbor-joining. Mol. Biol. Evol. 12:843–849.

HUELSENBECK, J. P., AND K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Syst. 28:437–466.

HUELSENBECK, J. P., AND B. RANNALA. 1997. Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. Science 276:227–232.

HUELSENBECK, J. P., B. RANNALA, AND Z. YANG. 1997. Statistical tests of host–parasite cospeciation. Evolution 51:410–419.

JUKES, T. H., AND C. R. CANTOR. 1969. Evolution of protein molecules. Pages 21–123 *in* Mammalian protein metabolism (H. N. Munro, ed.). Academic Press, New York.

KIDD, K. K., AND L. A. SGARAMELLA-ZONTA. 1971. Phylogenetic analysis: Concepts and methods. Am. J. Hum. Genet. 23:235–252.

KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide substitutions. J. Mol. Evol. 16:111–120.

KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequences, and the branching order of Hominoidea. J. Mol. Evol. 29:170–179.

KUHNER, M. K., AND J. FELSENSTEIN. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol. Biol. Evol. 11:459–468.

KUMAR, S., K. TAMURA, AND M. NEI. 1993. MEGA: Molecular evolutionary analysis, version 1.01. The Pennsylvania State Univ., University Park.

LEWIS, P. O. 1998. A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. Mol. Biol. Evol. 15:277–283.

LOCKHART, P., A. W. D. LARKUM, M. A. STEEL, P. J. WADDELL, AND D. PENNY. 1996. Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence evolution. Proc. Natl. Acad. Sci. USA 93:1930–1934.

LOCKHART, P. J., M. A. STEEL, A. C. BARBROOK, D. H. HUSON, M. A. CHARLESTON, AND C. J. LOWE. 1998. A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. Mol. Biol. Evol. 15:1183–1188.

MIYAMOTO, M. M., AND W. M. FITCH. 1995. Testing the covarion hypothesis of molecular evolution. Mol. Biol. Evol. 12:503–513.

MUSE, S. V., AND B. S. GAUT. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with applications to the chloroplast genome. Mol. Biol. Evol. 11:715–724.

NEI, M. 1991. Relative efficiencies of different tree-making methods for molecular data. Pages 90–128 *in* Phylogenetic analysis of DNA sequences (M. M. Miyamoto and J. Cracraft, eds). Oxford Univ. Press, Oxford.

OLMSTEAD, R. G., P. A. REEVES, AND A. C. YEN. 1998. Patterns of sequence evolution and implications for parsimony analysis of chloroplast DNA. Pages 164–187 *in* Molecular systematics of plants II: DNA sequencing (D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds.). Kluwer, Boston.

RAMBAUT, A., AND L. BROMHAM. 1998. Estimating divergence dates from molecular sequences. Mol. Biol. Evol. 15:442–448.

REYES, T. H., G. PESOLE, AND C. SACCONE. 1998. Complete mitochondrial DNA sequence of the fat dormouse *Glis glis*: Further evidence for rodent paraphyly. Mol. Biol. Evol. 15:499–505.

ROGERS, J. S. 1997. On the consistency of maximum likelihood estimation of phylogenetic trees from nucleotide sequences. Syst. Biol. 46:354–357.

ROGERS, J. S., AND D. L. SWOFFORD. 1998. A fast method for approximating maximum likelihoods of phylogenetic trees from nucleotide sequences. Syst. Biol. 47:77–89.

RZHETSKY, A., AND M. NEI. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. 9:945–967.

SHIMODAIRA, H., AND M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.

SIMON, C., F. FRATI, A. BECKENBACH, B. CRESPI, H. LIU, AND P. FLOOK. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87:651–701.

STRIMMER, K., AND A. VON HAESELER. 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13:964–969.

SULLIVAN, J., K. HOLSINGER, AND C. SIMON. 1995. Among-site rate variation and phylogenetic analysis of 12S rRNA in sigmodontine rodents. Mol. Biol. Evol. 12:988–1001.

SULLIVAN, J., K. E. HOLSINGER, AND C. SIMON. 1996. The effect of topology on estimates of among-site rate variation. J. Mol. Evol. 42:308–312.

SULLIVAN, J., J. A. MARKERT, AND C. WILLIAM KILPATRICK. 1997. Phylogeography and molecular systematics of the *Peromyscus aztecus* species group (Rodentia: Muridae) inferred using parsimony and likelihood. Syst. Biol. 46:426–440.

SULLIVAN, J., AND D. L. SWOFFORD. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mammal. Evol. 2:77–86.

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

SUNNUCKS, P., AND D. F. HALES. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). Mol. Biol. Evol. 13:510–524.

SWOFFORD, D. L. 1998. PAUP*. 1998. Phylogenetic analysis using parsimony (*and other methods). version 4. 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, 2nd edition (D. M. Hillis, C. Motitz, and B. K. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.

SZYMURA, J. M., D. H. LUNT, AND G. M. HEWITT. 1996. The sequence and structure of the meadow grasshopper (*Chorthippus parallelus*) mitochondrial srRNA, ND2, COI, COII, ATPase8 and 9 tRNA genes. Insect Mol. Biol. 5:127–139.

TAKEZAKI, N., AND T. GOJOBORI. 1999. Correct and incorrect vertebrate phylogenies obtained by entire mitochondrial DNA sequences. Mol. Biol. Evol. 16: 590–601.

TATENO, Y., N. TAKEZAKI, AND M. NEI. 1994. Relative efficiencies of the maximum-likelihood, neighbor-joining and maximum parsimony methods when substitution rate varies with site. Mol. Biol. Evol. 11: 261–277.

TUFFLEY, C., AND M. STEEL. 1998. Modeling the covarion hypothesis of nucleotide substitution. Math. Biosci. 147:63–91.

VAN DE PEER, Y., J.-M. NEEFS, P. DE RIJK, AND R. DE WACHTER. 1993. Reconstructing relationships from eukaryotic small-ribosomal subunit RNA sequences: Calibration of the molecular clock. J. Mol. Evol. 37:221–232.

VOELKER, G., AND S. V. EDWARDS. 1998. Can weighting improve bushy trees? Models of cytochrome *b* evolution and the molecular systematics of pipits and wagtails (Aves: Motacillidae). Syst. Biol. 47:589–603.

WADDELL, P. J., Y. CAO, J. HAUF, AND M. HASEGAWA. 1999. Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid–invariant sites—LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. Syst. Biol. 48:31–53.

WADDELL, P. J., D. PENNY, AND T. MOORE. 1997. Hadamard conjugations and modeling sequence evolution with unequal rates across sites. Mol. Phylogenet. Evol. 8:33–50.

WADDELL, P. J., AND M. A. STEEL. 1997. General time-reversible distances with unequal rates across sites: Mixing Γ and inverse Gaussian distributions with invariant sites. Mol. Phylogenet. Evol. 8:398–414.

WAKELEY, J. 1994. Substitution-rate variation and the estimation of transition bias. Mol. Biol. Evol. 11:436–442.

WAKELEY, J. 1996. The excess of transitions among nucleotide substitutions: New methods of estimating transition bias underscore its significance. Trends Ecol. Ecol. 11:158–163.

WALD, A. 1949. Note on the consistency of the maximum likelihood estimate. Ann. Math. Stat. 20:595–601.

WHELAN, S., AND N. GOLDMAN. 1999. Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. Mol. Biol. Evol. 16:1292–1299.

YANG, Z. 1993. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10:1396–1401.

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. 1994c. Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. Syst. Biol. 43:329–342.

YANG, Z. 1995. A space time process model for the evolution of DNA sequences. Genetics 139:993–1005.

YANG, Z. 1996a. Phylogenetic analysis using parsimony and likelihood methods. J. Mol. Evol. 42:294–307.

YANG, Z. 1996b. Among-site rate variation and its impact on phylogenetic analysis. Trends Ecol. Evol. 11:367–372.

YANG, Z. 1996c. Maximum likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42:587–596.

YANG, Z. 1997. PAML: A program for phylogenetic analysis by maximum likelihood. CABIOS 13:555–556.

YANG, Z., N. GOLDMAN, AND A. E. FRIDAY. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood estimation. Mol. Biol. Evol. 11:316–324.

YANG, Z., N. GOLDMAN, AND A. E. FRIDAY. 1995. Maximum likelihood trees from DNA sequences: A peculiar statistical estimation problem. Syst. Biol. 44:384–399.

YANG, Z., AND A. D. YODDER. 1999. Estimation of the transition/transversion rate bias and species sampling. J. Mol. Evol. 48:274–283.

ZHANG, J. 1999. Performance of likelihood ratio tests of evolutionary hypotheses under inadequate substitution models. Mol. Biol. Evol. 16:868–875.

*Received 22 September 1999; accepted 19 January 2000*
*Associate Editor: K. Crandall*