The Effects of Nucleotide Substitution Model Assumptions on Estimates of Nonparametric Bootstrap Support

Thomas R. Buckley and Clifford W. Cunningham
Department of Biology, Duke University, Durham, North Carolina

The use of parameter-rich substitution models in molecular phylogenetics has been criticized on the basis that these models can cause a reduction both in accuracy and in the ability to discriminate among competing topologies. We have explored the relationship between nucleotide substitution model complexity and nonparametric bootstrap support under maximum likelihood (ML) for six data sets for which the true relationships are known with a high degree of certainty. We also performed equally weighted maximum parsimony analyses in order to assess the effects of ignoring branch length information during tree selection. We observed that maximum parsimony gave the lowest mean estimate of bootstrap support for the correct set of nodes relative to the ML models for every data set except one. For several data sets, we established that the exact distribution used to model among-site rate variation was critical for a successful phylogenetic analysis. Site-specific rate models were shown to perform very poorly relative to gamma and invariables sites models for several of the data sets most likely because of the gross underestimation of branch lengths. The invariable sites model also performed poorly for several data sets where this model had a poor fit to the data, suggesting that addition of the gamma distribution can be critical. Estimates of bootstrap support for the correct nodes often increased under gamma and invariable sites models relative to equal rates models. Our observations are contrary to the prediction that such models cause reduced confidence in phylogenetic hypotheses. Our results raise several issues regarding the process of model selection, and we briefly discuss model selection uncertainty and the role of sensitivity analyses in molecular phylogenetics.

Introduction

Model-based methods of phylogenetic analysis are becoming increasingly popular because of the inherent ability to account for properties of the evolutionary process during tree selection. The importance of evolutionary models has been underscored by a number of recent studies (e.g., Lockhart et al. 1996; Sullivan and Swofford 1997), which have shown that unrealistic assumptions concerning the evolutionary process can cause systematic error. Ignoring the major features of the evolutionary process can cause misinterpretation of the data and subsequent errors in phylogenetic reconstruction. Furthermore, if enough data are available, misspecified models can lead to very strong statistical support for incorrect hypotheses (Yang, Goldman, and Friday 1995), as estimated by nonparametric bootstrapping (Efron 1979; Felsenstein 1985).

In molecular phylogenetics, a nonparametric bootstrap analysis involves the resampling of nucleotide or amino acid sites from an alignment, with replacement, so as to generate a number of psuedoreplicate data sets of the same length as the original alignment. The optimal topology or any other quantity of interest is then estimated from each pseudoreplicate. Bootstrapping is generally regarded as a method for estimating the variance associated with any particular parameter for which the sampling distribution is not obvious (Manly 1997). However, the exact statistical interpretation of bootstrap proportions on a phylogenetic tree has been under debate for some time (Sanderson 1989, 1995; Felsenstein and Kishino 1993; Hillis and Bull 1993; Efron, Hallor-

an, and Holmes 1996; Swofford et al. 1996; Andrieu, Caraux, and Gascuel 1997; Lee 2000). Most authors, at least implicitly, interpret bootstrap proportions as rough measures of statistical support for a node, given the data and the assumptions of the phylogenetic method.

Under model-based methods of phylogenetic estimation, bootstrap proportions can vary greatly as the parameter richness of the model changes (Sullivan, Markert, and Kilpatrick 1997; Waddell and Steel 1997; Cunningham, Zhu, and Hillis 1998; Krajewski et al. 1999; Tourasse and Guoy 1999; Inagaki et al. 2000; Buckley, Simon, and Chambers 2001). For example, Buckley, Simon, and Chambers (2001) observed that the addition of among-site rate variation parameters to the substitution model caused the bootstrap support for some nodes to increase and others to decrease, relative to bootstrap proportions estimated under the assumption of equal rates among sites. It is commonly assumed that changes in bootstrap proportions associated with shifts in model structure are caused by the counteracting forces of variance and bias. In the context of phylogenetics, bias refers to the failure of an estimator to accurately reconstruct the true topology because of misleading assumptions (e.g., model misspecification). In the case of tree topology, variance would be problematic to quantify but is an inevitable outcome of increasing model complexity.

Given the bias-variance relationship, there are four possible explanations for changes in bootstrap support associated with changes in model complexity. First, as a model becomes increasingly parameterized, bootstrap values may decrease because the sampling variance associated with the model increases (Yang, Goldman, and Friday 1995). Second, bootstrap values may also increase because the improving fit between the model and the data boosts phylogenetic accuracy. Third, if the extraneous parameters are grossly unrealistic, then the bias

of the model may actually increase, causing a drop in accuracy over pseudoreplicates. Fourth, the addition of parameters to a model may have little or no effect on either bias or variance, or alternatively any decrease in bias may be negated by a corresponding increase in variance. In this situation, we would expect our confidence in the estimated phylogeny to remain unchanged. Because of the counteracting forces of variance and bias it is difficult to predict the effects on bootstrap proportions as a model becomes increasingly parameterized, especially when the models under consideration differ by only a few parameters. Objective model selection is an attempt to choose a model that avoids the dual problems of both under- and overfitting (Akaike 1973; Goldman 1993; Frati et al. 1997; Burnham and Anderson 1998; Posada and Crandall 1998; Forster 2000; Wasserman 2000).

These predictions and observations conflict with the claims of some authors (e.g., Reyes, Pesole, and Saccone 1998, 2000; Takahashi and Nei 2000; Xia 2000) who have questioned the utility of complex models in phylogenetic analysis, and in particular, models that assume unequal substitution rates among sites. Furthermore, some authors have also questioned the appropriateness (Takahashi and Nei 2000) or practicality (Sanderson and Kim 2000) of using objective model selection criteria for choosing a substitution model.

In the following study, we explored the effects of substitution model assumptions on bootstrap proportions by analyzing a variety of data sets for which all the evolutionary relationships are known with a very high degree of confidence. We also compared the results from equally weighted parsimony analyses because this method makes a stringent set of assumptions, such as equal substitution rates among sites and between nucleotides (Yang 1996) that can be relaxed using ML methods. Furthermore, the parsimony method ignores branch length information during tree selection, unlike the commonly used ML methods. Our results strongly contradict the argument that complex substitution models are only useful for extremely long sequences (e.g., Takahashi and Nei 2000) and also have implications for the process of model selection in molecular phylogenetics.

Materials and Methods
Data Sets
Bird Whole Mitochondrial Genomes

We analyzed sequences from whole mitochondrial genomes (14,043 bp from all tRNA-, rRNA-, and protein-coding genes, with the exception of ND6) from nine taxa including turtle, alligator, rhea, ostrich, falcon, duck, oscine songbird, and suboscine songbird (Mindell et al. 1999). Relationships among these species are known with a high degree of confidence based on morphological characters (Cracraft 2001), DNA-DNA hybridization data (Sibley and Ahlquist 1990), immunological data (Prager and Wilson 1980), various nuclear genes (Stapel et al. 1984; Groth and Barrowclough 1999; Garcia-Moreno and Mindell 2000; Lovette and

Bermingham 2000), and mitochondrial and nuclear rRNA genes (van Tuinen, Sibley, and Hedges 2000).

Arthropod EF1α

Phylogenetic analyses of a wide variety of gene sequences all support the hypothesis that hexapods and crustaceans form a monophyletic group (Pancrustacea) to the exclusion of myriapods and chelicerates (Friedrich and Tautz 1995, 2000; Regier and Schultz 1997, 2001; Boore, Lavrov, and Brown 1998; García-Monchado et al. 1999; Schultz and Regier 2000; Wilson et al. 2000; Giribet, Edgecombe, and Wheeler 2001; Hwang et al. 2001). We have reanalyzed eight EF1α sequences (1,092 bp) from Regier and Schultz (1997) and Schultz and Regier (2000) including; *Periplaneta americana* (cockroach), *Pedetontus saltator* (Archaeognatha insect), *Tomocerus* sp. (Collembola), *Nebalia hessleri* (Malacostracan crustacean), *Armadillidium vulgare* (Malacostracan crustacean), *Narceus americanus* (Myriapoda), *Scutigera coleoptrata* (Myriapoda), *Dysdera crocata* (Spider), and *Dinothrombium pandorae* (Spider).

Primate Mitochondrial DNA

We obtained the alignment of 888 bp of mitochondrial DNA (mtDNA) sequences (protein-coding and tRNA genes) from nine primate species, distributed with the PAML package (Yang 1997a) and originally described by Hayasaka, Gojobori, and Horai (1988). Most systematists recognize that human and chimpanzee are sister species, based on analyses of a wide variety of molecular markers (e.g., Shoshani et al. 1996; Satta, Klein, and Takahata 2000). The relationships among the other primate species are corroborated by whole mitochondrial sequences (Waddell et al. 1999), Alu repeats (Zietkiewicz, Richer, and Labuda 1999), various nuclear genes (Goodman et al. 1998; Murphy et al. 2001; Page and Goodman 2001), morphological, and fossil data (Goodman et al. 1998).

Insect 18S rRNA

We analyzed a subset of the insect 18S rRNA sequences (902 bp) from Whiting et al. (1997), also analyzed by Huelsenbeck (1998), including a collembolan (*Hypogastura* sp.), an odonate (*Libellula pulchella*), a plecopteran (*Cultus decisus*), a lepidopteran (*Ascalapha odorata*), a trichopteran (*Pycnopsyche lepida*), a mecopteran (*Bittacus chlorostigmus*), and a siphonapteran (*Ctenocephalides canis*). The relationships among all these seven taxa are known with a high degree of confidence (Kristensen 1995; Whiting et al. 1997).

Vertebrate Cytochrome Oxidase Subunit II

Russo, Takezaki, and Nei (1996) examined the performance of a range of phylogenetic methods using different mitochondrial genes from a group of vertebrates (carp, loach, trout, frog, chicken, opossum, mouse, rat, cow, blue whale, and finback whale) where all of the relationships are well known. These authors concluded

Table 1
AIC Differences for each Data Set Analyzed

	Bird mtDNA	Arthropod EF1α	Primate mtDNA	Insect 18S rRNA	Vertebrate COII	Rodent mtDNA
JC69 (0)	17,115.076	1,578.818	1,193.316	233.325	1,451.106	518.072
F81 (3)	14,639.222	1,549.349	935.571	234.679	1,301.903	406.634
HKY85 (4)	12,132.099	1,338.093	523.355	173.897	1,068.698	270.228
GTR (8)	9,393.294	1,227.198	444.576	148.050	966.911	224.015
GTR + I (9)	2,188.336	364.700	174.692	6.187	445.880	109.360
GTR + Γ (9)	807.290	346.311	124.561	Best	354.934	109.119
GTR + I + Γ (10)	799.881	318.725	125.933	2.000	343.730	111.080
GTR + SSR (10, 11)	Best	Best	Best	—	Best	Best

NOTE.—The number of free parameters are given in parentheses after each model. The two values given after the SSR model represent the number of parameters when only protein-coding sites are included (three rate classes) and when RNA coding sites (four rate classes) are included, respectively.

that one of the worst-performing genes was the cytochrome oxidase subunit II gene (COII, 681 bp).

Sigmodontine Rodents Mitochondrial DNA

Sullivan, Holsinger, and Simon (1995) analyzed mitochondrial 12S rRNA and cytochrome *b* sequences from a group of sigmodontine rodents for which the relationships among taxa were known with a high degree of confidence. Although a combined analysis of both genes yielded the well-supported phylogeny, the 12S rRNA gene, when analyzed separately, yielded a strongly supported and incorrect topology. These taxa are all very closely related with the oldest divergence estimated to have occurred approximately 6 MYA (Sullivan, Holsinger, and Simon 1995).

Nucleotide Substitution Models and Nonparametric Bootstrapping Strategy

All analyses were done using various beta versions of PAUP* 4.0 (Swofford 1998). We restricted our analyses to equal-weighted maximum parsimony (Fitch 1971) and eight different ML nucleotide substitution models, including JC69 (Jukes and Cantor 1969), F81 (Felsenstein 1981), HKY85 (Hasegawa, Kishino, and Yano 1985), GTR (e.g., Yang 1994*a*), GTR + I (invariable sites; Hasegawa, Kishino, and Yano 1985), GTR + Γ (gamma-distributed rates; Yang 1994*b*), GTR + I + Γ (mixed invariable sites and gamma-distributed rates; Gu, Fu, and Li 1995), and GTR + SSR (site-specific rates; Swofford et al. 1996). For the SSR models, we partitioned the data into the three codon positions and pooled any tRNA or rRNA sites that were present into a single rate category. We used eight rate categories for the gamma distribution and ML estimates of base frequencies.

We evaluated the fit of these models to the data using the Akaike information criterion (AIC; Akaike 1973). We chose the AIC for model selection rather than the more commonly used likelihood ratio tests because the AIC allows non-nested models to be ranked and compared (e.g., SSR and Γ-rates models), and facilitates the identification of groups of models that have similar fits to the data (see Burnham and Anderson 1998). Following Burnham and Anderson (1998), we present AIC differences (Δ_i). The AIC difference for model *i* is: Δ_i = AIC$_i$ − min AIC, where min AIC is the best-fit model. AIC values were calculated from the well-supported topologies for each data set. Tree lengths were estimated under each substitution model by summing the branch lengths for all branches in the true tree.

We used a variety of search strategies depending on the computational burden of analyzing each data set. For the primate mtDNA, arthropod EF1α, vertebrate COII, and sigmodontine rodent mtDNA sequences, we obtained starting trees using stepwise addition with 10 random addition replicates and analyzed 500–1,000, 500, 500, and 2000 bootstrap replicates, respectively. For the bird mtDNA and insect 18S rRNA data sets, we analyzed 100–200 and 1,000 replicates, respectively, with neighbor-joining (Saitou and Nei 1987) start trees. For all data sets, we employed TBR branch swapping. All the parsimony analyses consisted of 2,000 bootstrap replicates with start trees obtained by stepwise addition with 10 random addition replicates. For each data set, we estimated model parameters from the true tree and then fixed those parameter values for each replicate. For some of the data sets, we repeated the analyses using different search strategies and determined that varying the number of replicates or the method of obtaining a start tree had little effect on bootstrap proportions for these data.

To generate pseudoreplicate data sets for the GTR + SSR analyses, we used CodonBootstrap 2.21 (J. Bollback, personal communication), which samples codons rather than individual nucleotides. This process allows the coding structure of the gene to be maintained over different pseudoreplicates (Bollback and Huelsenbeck 2001).

Results

For each data set, we present the results from the AIC model fitting (table 1), parameter estimates under the GTR + Γ model (table 2), tree lengths (table 3), and mean bootstrap proportions for the set of correct nodes (table 4). We discuss later, the influence of substitution model assumptions on estimates of bootstrap support for each data set.

Table 2
Parameter Estimates for Each Data Set Under the GTR + Γ Model

	Bird mtDNA	Arthropod EF1α	Primate mtDNA	Insect 18S rRNA	Vertebrate COII	Rodent mtDNA
π_A.....	0.300	0.281	0.347	0.262	0.351	0.365
π_C.....	0.310	0.231	0.312	0.218	0.286	0.223
π_G.....	0.151	0.239	0.090	0.278	0.114	0.156
π_T.....	0.239	0.248	0.251	0.242	0.249	0.256
r_{AC}	4.574	2.623	12.221	1.175	0.727	6.447
r_{AG}	7.377	3.409	79.557	1.656	4.342	13.859
r_{AT}.....	1.415	2.428	8.204	1.046	1.323	6.867
r_{CG}	0.383	1.603	7.302	0.393	0.396	4.167
r_{CT}.....	8.380	12.011	83.295	4.430	8.843	46.888
α......	0.394	0.332	0.423	0.273	0.274	0.137

NOTE.—The r_{GT} parameter is scaled to 1.0 for all data sets.

Bird Whole Mitochondrial Genomes

For this data set, the well-supported topology was only recovered under the GTR + Γ and GTR + I + Γ models. Our bootstrap analyses (fig. 2*A*) show that support for the correct root of the avian tree (node A) is moderate under the GTR + Γ (64%) and GTR + I + Γ (71%) models, and is extremely low under the equal rates models (1% to 6%). The GTR + I model has a poor fit to the data relative to the GTR + Γ, GTR + I + Γ, and GTR + SSR₄ models (table 1), and yields low support for node A (36%). The GTR + SSR₄ model has the best fit to the data (table 1), yet this model gives only 15% bootstrap support for node A (fig. 2*A*). Additionally, the GTR + I and GTR + SSR₄ models also gave lower estimates of the total tree length relative to the GTR + Γ and GTR + I + Γ models (table 3). The GTR + I + Γ model gave the highest mean bootstrap proportion for the set of correct nodes (table 4).

Under the equal rates, GTR + I, and GTR + SSR₄ models, nodes B and E are disrupted by the outgroup taxa rooting the tree within the song birds. Accordingly, these two nodes also received much higher support under the GTR + Γ and GTR + I + Γ models. Both nodes C and D were very well supported, with the GTR + Γ and GTR + I + Γ models, giving the highest bootstrap values for node C. All models gave 100% bootstrap support for node D.

Like the simple ML models, the parsimony analysis performed very poorly in recovering the true tree across bootstrap replicates. For example, nodes A, B, and E were never observed in any of the 2,000 replicates. The parsimony method also gave the lowest mean bootstrap proportion for the correct nodes.

Arthropod EF1α

The correct tree was recovered only under the GTR + I, GTR + Γ, and GTR + I + Γ models for these data. For three of the nodes in the arthropod EF1α phylogeny (fig. 1*B*), we observed large differences in support among the various models (fig. 2*B*). Node C, the monophyly of the Pancrustacea, was poorly supported under parsimony (26%) and under the JC69 model (49%), whereas support peaked under the GTR + I model (77%). Support for node E (hexapod monophyly) was highest under the among-site rate variation models, especially under the GTR + Γ model (54%). Support for node F was lowest under the GTR + SSR₃ model (47%) and highest under the GTR + I model (70%). For this data, the GTR + I model indicated the highest support for the set of nodes in the true tree (table 4), although this model gave lower estimates of the total tree length relative to the GTR + Γ and GTR + I + Γ models (table 3). Furthermore, support for the GTR + I model was extremely low under the AIC, even when the GTR + SSR₃ model was excluded from the set of candidate models (table 1). Parsimony yielded the lowest mean bootstrap proportion for the set of correct nodes (table 4).

Table 3
Varied Sites and Tree Lengths Estimated Under Different Substitution Models

	Bird mtDNA	Arthropod EF1α	Primate mtDNA	Insect 18S rRNA	Vertebrate COII	Rodent mtDNA
No. of sites........	14,043	1,092	888	902	681	1,028
% Varied sites	52.8	49.5	56.3	24.6	56.2	18.1
JC69	1.383	1.481	1.285	0.461	1.664	0.252
F81	1.389	1.476	1.296	0.461	1.663	0.252
HKY85...........	1.409	1.492	1.361	0.464	1.717	0.255
GTR	1.391	1.499	1.349	0.467	1.690	0.256
GTR + I..........	1.770	1.864	1.708	0.574	2.168	0.305
GTR + Γ	2.584	**3.000**	2.752	0.605	3.740	0.309
GTR + I + Γ	**2.632**	2.380	**2.682**	**0.616**	**4.015**	**0.310**
GTR + SSR.......	1.750	1.821	1.713	—	2.450	0.284

Table 4
Mean Bootstrap Support Values on the True Trees Under Various Substitution Models

	Bird mtDNA	Arthropod EF1α	Primate mtDNA	Insect 18S rRNA	Vertebrate COII	Rodent mtDNA
Parsimony........	35.2	61.3	86.3	66.8	76.5	89.8
JC69	38.3	68.2	88.3	70.5	80.5	89.2
F81	38.6	68.5	88.7	70.8	80.6	89.6
HKY85..........	40.4	70.7	91.0	74.0	86.9	**92.2**
GTR	41.6	72.5	91.8	73.8	90.5	92.0
GTR + I.........	60.8	**76.5**	91.0	70.8	**92.9**	90.4
GTR + Γ.........	75.2	72.8	95.5	74.3	87.6	89.6
GTR + I + Γ	**78.2**	75.5	**96.0**	**74.7**	86.9	89.2
GTR + SSR......	48.0	71.0	90.5	—	89.1	78.4

Primate Mitochondrial DNA

Parsimony and the JC69 and F81 models led to the selection of an incorrect topology for the primate data set. The remaining models all converged on the correct topology. For the primate mtDNA data set, we observed that all nodes, with the exception of the human-chimpanzee node (A), were very well supported, irrespective of the substitution model. The strongest bootstrap support values for node A were obtained under the parameter-rich GTR + Γ (86%) and GTR + I + Γ (85%) models (fig. 2C). The equal rates models gave much lower estimates of bootstrap support ranging from 44% (JC69) to 58% (GTR). The performance of parsimony for this node was also poor (47%). The model with the lowest AIC value for this data set, and therefore the best-fit model, was GTR + SSR$_4$ (table 1), which gave a bootstrap support value of only 51% for the chimpanzee-human node (fig. 2C). Of all the among-site rate variation models employed, the GTR + I model had the worst fit to the data (table 1) and gave a bootstrap value of only 56% for the chimpanzee-human node. The estimated total length of the tree was the highest under the GTR + I + Γ model (table 3), and additionally, this model yielded the highest mean bootstrap value for the set of correct nodes, whereas parsimony gave the lowest mean bootstrap proportion (table 4).

Insect 18S rRNA

Only the GTR + Γ and GTR + I + Γ models led to the selection of the correct topology for these data. Bootstrap support for node A (fig. 2D) is highest under the among-site rate variation models, ranging from 52% (GTR + I) to 59% (GTR + I + Γ). In many of the bootstrap analyses, node A is disrupted because the long plecopteran branch joins with the long collembolan branch, and this attraction is reduced under the more complex models. As with all the other data sets, the GTR + I model has the worst fit to the data of the among-site rate variation models (table 1) and gives slightly lower estimates of bootstrap support for node A.

Both the lepidopteran and the trichopteran branches are relatively long and share a common node (C), a different situation from the plecopteran and the collembolan branches. Based on previous simulation studies (e.g., Yang 1997b), we expected a priori that the equal rates ML models would yield greater support for this node

than the among-site rate variation models, and this is indeed what we observed. This node receives very strong support under the equal rates models (97% to 99%) and parsimony (94%), whereas support is slightly lower under the among-site rate variation models (87% to 89%). Node B, supporting the monophyly of the Holometabola, is strongly supported under ML (98% to 99%) but slightly less so under parsimony (88%). The grouping of the short siphonapteran and mecopteran branches (node D) received low bootstrap support under all models, with the support from the among-site rate variation models being slightly lower than that from the equal rates models (fig. 2D).

The GTR + I + Γ model gave the highest mean bootstrap proportion, although this value was very similar to that from a number of other models (table 4). Parsimony gave the lowest mean bootstrap proportion. The GTR + I + Γ model also has the highest estimate of the total length of the tree (table 3).

Vertebrate Cytochrome Oxidase Subunit II

With the single exception of node E, most nodes in this data set were relatively easy to reconstruct. Additionally, all methods and models with the exception of parsimony and ML under the JC69 and F81 models led to selection of the correct topology. Support for node E was lowest under the JC69 model (16%) and increased steadily as the substitution model became more complex, and peaked under the GTR + SSR$_3$ model (84%). Node F also received much greater support under the among-site rate variation models than under the equal rates models and parsimony. Bootstrap values for nodes C and G were more erratic with respect to model complexity. The GTR + I model inferred the largest mean bootstrap proportion for the correct nodes, whereas parsimony yielded the lowest mean bootstrap proportion.

Sigmodontine Rodents Mitochondrial DNA

All models led to the selection of the correct topology for these data. As with the other data sets, the GTR + SSR$_4$ model had the lowest AIC value (table 1). There was only moderate variation in bootstrap support among the different models (fig. 2F). However, support for node B was much lower under the GTR + SSR$_4$ model (51%) than under parsimony (92%) or the remaining ML models (78% to 92%). The rodent mtDNA
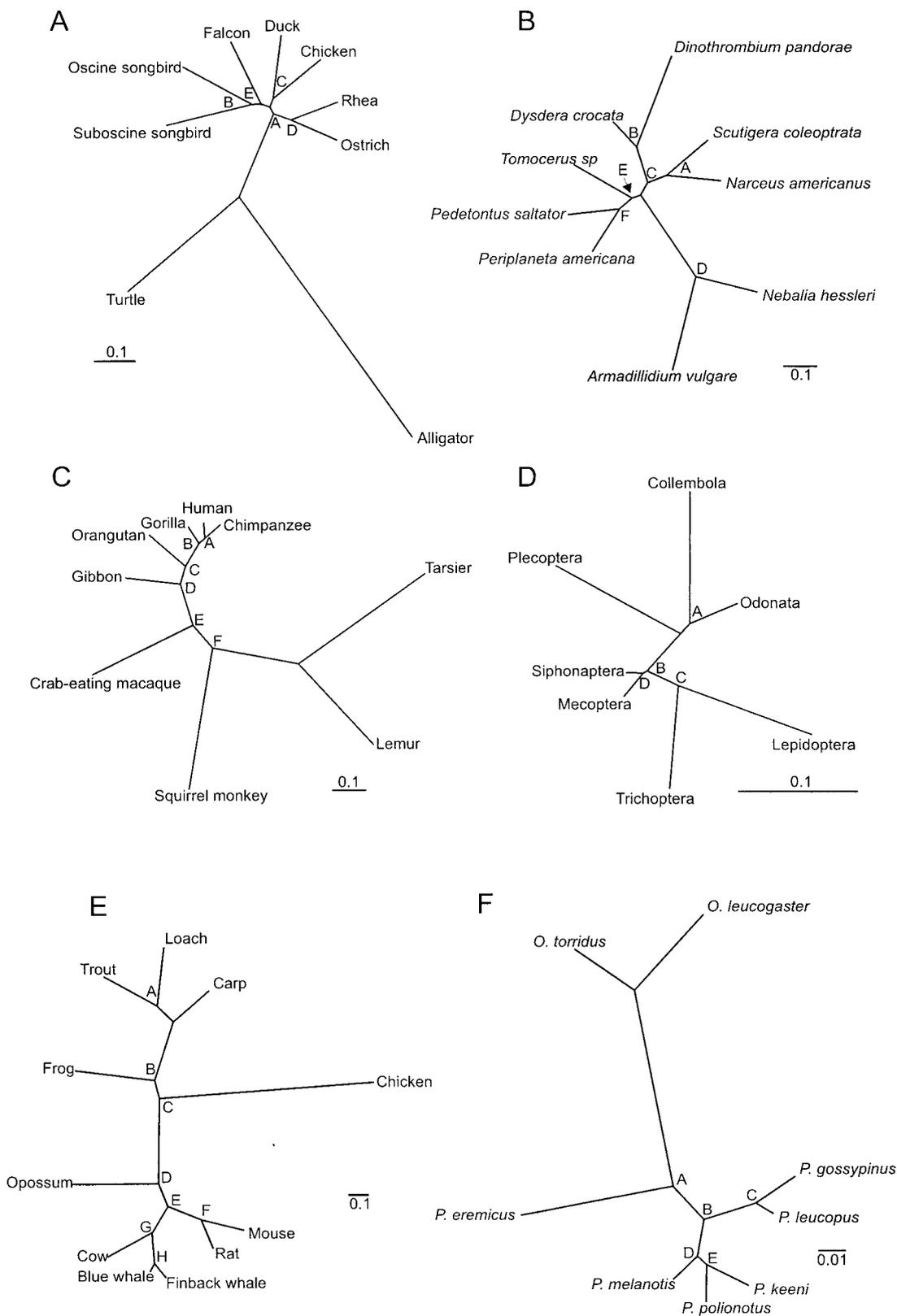
FIG. 1.—Well-established phylogenetic relationships for the taxa analyzed in this study: (*A*) Bird whole mitochondrial genomes; (*B*) arthropod EF1α sequences; (*C*) primate mtDNA sequences; (*D*) insect 18S rRNA sequences; (*E*) vertebrate COII sequences; and (*F*) sigmodontine rodent mtDNA sequences. The branch lengths were estimated under the GTR + Γ model.
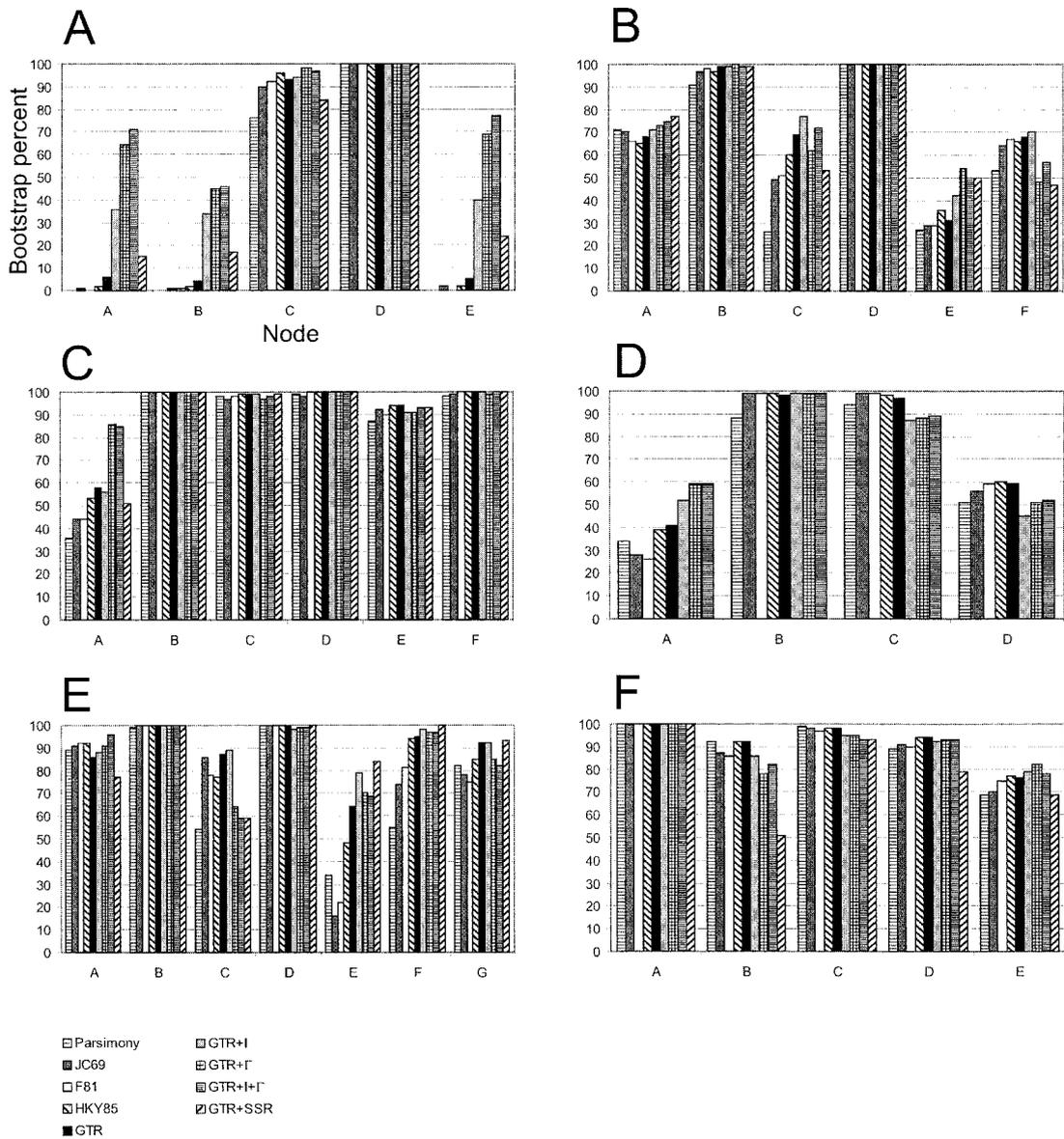
FIG. 2.—Graphs showing bootstrap values estimated under parsimony and the nine nucleotide substitution models for each correct node. The nodes are labeled as in figure 1. (*A*) Bird whole mitochondrial genomes, (*B*) arthropod EF1α sequences, (*C*) primate mtDNA sequences, (*D*) insect 18S rRNA sequences, (*E*) vertebrate COII sequences, and (*F*) sigmodontine rodent mtDNA sequences.

was the only data set where an equal rates model (HKY85) yielded the highest mean bootstrap support for the correct nodes (table 4). The GTR + I + Γ model had the lowest mean bootstrap proportion (table 4). The GTR + I, GTR + Γ, and GTR + I + Γ models gave very similar estimates of the total tree length (table 3) and additionally had similar AIC differences (table 1).

Discussion
Maximum Parsimony Versus Maximum Likelihood

Drawing general conclusions from a small number of data sets can be problematic because we do not know how representative the example data sets are of typical phylogenetic problems. The parameter estimates (table 2) and tree lengths (table 3) are typical of many data sets, and therefore our results should have more general

implications. The known phylogeny approach (e.g., Cunningham, Zhu, and Hillis 1998) does have an advantage over simulation studies of phylogenetic methods (e.g., Huelsenbeck and Hillis 1993) because the former avoids the simplifying assumptions inherent in generating data under a known model. The approach of using strongly corroborated empirical phylogenies represents an alternative approach to evaluating phylogenetic methods (e.g., Allard and Miyamoto 1992; Sullivan, Holsinger, and Simon 1995; Cunningham 1997; Naylor and Brown 1998). We view the approach that we have taken here as complementary to simulation studies and studies of artificially generated phylogenies.

Of all the phylogenetic methods that we investigated, maximum parsimony yielded the lowest mean bootstrap proportion for every data set with the excep-

tion of the rodent mtDNA data. These results agree with theoretical studies (e.g., Yang 1996; Swofford et al. 2001) that have demonstrated the utility of ML coupled with realistic models when rates of change vary among branches (e.g., the bird mtDNA and insect 18S rRNA; fig. 1). The increased statistical support for the correct relationships that we observed under ML also likely stems from the increased accuracy of this method relative to equally weighted parsimony because of its inherent ability to accommodate biases in the data, such as among-site rate variation. We also observed that the JC69 model always outperformed parsimony except for the rodent mtDNA, in agreement with the simulation study of Yang (1996). Parsimony did indicate slightly higher support for the set of correct nodes for the rodent mtDNA data set, which we attribute to the inflation of variance associated with a complex model coupled with a small number of varied sites (table 3). However, it is important to note that the more complex ML models did not indicate strong statistical support for incorrect hypotheses.

Single-rate Versus Among-site Rate Variation Models

Some authors have questioned the utility of substitution models that explicitly describe the distribution of among-site rate variation. For example, based on the results of a simulation study, Takahashi and Nei (2000) claimed that the HKY85 and gamma models are only useful when extremely long sequences are available (>10,000 bp). Reyes, Pesole, and Saccone (2000) asserted that if an among-site rate variation model does not match the process that generated the data, then the results obtained under the more complex model can be less significant than results obtained under an equal rates model. The results presented here provide a number of empirical refutations to these arguments. Our results have clearly illustrated that the relationship between the parameter-richness of the assumed substitution model and the confidence assigned to a phylogenetic hypothesis is complex, as one would expect, as a result of the counteracting forces of variance and bias. Increases in bootstrap support resulting from more accurate modeling of the substitution process presumably results from a decrease in bias. Eventually, we expect that bootstrap values would begin to decrease as the model becomes overparameterized. Our results suggest that this point has not been reached for the combination of models and data sets analyzed here. The use of among-site rate variation models certainly does not necessarily lead to a decrease in accuracy or statistical support for phylogenetic hypotheses and, in most of the situations described here, does quite the opposite.

For every data set that we analyzed, with the exception of the rodent mtDNA sequences, the mean bootstrap proportion for the correct nodes was highest for one of the among-site rate variation models. Even for sequences as short as the primate mtDNA data set (888 bp), the parameter-rich GTR + I + Γ model yielded the highest mean bootstrap value for the correct tree. Some of the data sets showed only small differences in mean bootstrap support among models. On this basis, it could be argued that the less complex and more computationally efficient models work just as well. However, within such data sets, bootstrap proportions for some individual nodes exhibited relatively large shifts in support. For example, the greatest difference in mean bootstrap support between models for the primate data set is 9.7% (parsimony vs. GTR + I + Γ). However, the support for some nodes was less stable than that suggested by this figure. For example, bootstrap support for node A varied from 36% (parsimony) to 86% (GTR + Γ).

As with the other data sets, support for some nodes in the rodent mtDNA phylogeny increased with the inclusion of among-site rate variation parameters, and the support for others decreased. However, for this data set, the highest mean bootstrap proportion was obtained under the equal rates HKY85 model. The overall reduced support for the true tree in the rodent data set under the among-site rate variation models agrees with the simulation studies of Takahashi and Nei (2000) who observed a slight reduction in phylogenetic accuracy when rates across sites models were used to estimate trees from sequences with a very small number of variable sites. However, we dispute the general conclusions of these authors that among-site rate variation models have little utility in general because these models worked very well for the other data sets that we analyzed here, including short sequences (e.g., insect 18S rRNA and vertebrate COII sequences) and sequences from closely related species (e.g., the human-chimpanzee-gorilla radiation). It can be instructive to compare the results from simple models (e.g., JC69) with those of more complex models for data sets of closely related taxa as advocated by Stanger-Hall and Cunningham (1998). However, we do not advocate running simple models exclusively unless this can be justified statistically and biologically.

The pattern of variation in bootstrap support values among the various models that we observed agrees with the increasingly well-known and predictable behavior of various phylogenetic methods in different regions of tree space (Felsenstein 1978; Yang 1996, 1997b; Cunningham, Zhu, and Hillis 1998; Huelsenbeck 1998; Siddall 1998; Bruno and Halpern 1999; Steel and Penny 2000; Swofford et al. 2001). We observed a potential example of long-branch attraction in the insect 18S rRNA sequences where support for node A is very low under parsimony and the equal rates ML models because of an attraction between the long plecopteran and the long collembolan branches. Conversely, in this same data set, we also observed reduced support under the among-site rate variation models relative to parsimony and the equal rates models for node C where two long branches (Lepidoptera and Trichoptera) are correctly joined. This observation probably results from misinterpretation of the data by the underfitted ML models and parsimony because of the unequal branch lengths (Felsenstein 1978; Yang 1997b; Huelsenbeck 1998; Bruno and Halpern 1999). This observation also supports the theoretical predictions (e.g., Bruno and Halpern 1999; Swofford et al. 2001) that best-fit ML models can require more data than parsimony or an underfitted ML model to achieve

equivalent levels of statistical support for such nodes. However, we agree with Swofford et al. (2001) that the realistic ML models give more appropriate estimates of confidence for such nodes. Contrary to the expectations of Siddall (1998) and Siddall and Whiting (1999), we did not observe an example in any of the data sets that we have analyzed here where a complex model yields drastically higher estimates of statistical support for an incorrect node when parsimony or an underfitted ML converged on the correct node.

Our results suggest that the inflation of variance associated with among-site rate variation models (e.g., Reyes, Pesole, and Saccone 2000; Takahashi and Nei 2000) is not a major concern for these data sets, probably because the increase in variance is often offset by a decrease in bias. Our results exemplify the general point made by Shibata (1989) that model underfitting is potentially a much worse problem than overfitting. It will be interesting to see whether our observations are consistent with those from further empirical studies.

Site-specific Rate Models Versus Gamma and Invariable Sites Models

We observed that for several data sets, the SSR models seriously misled the analyses, even when these models were preferred by the AIC. For example, the AIC indicated that the GTR + SSR$_4$ model is the best-fit model for the bird mtDNA data set. However, this model was incapable of recovering the true phylogeny for this data set, and worse, it provided extremely low support for the correct root of the avian tree (15%). In contrast, the AIC indicated that the GTR + I + Γ model fitted the data poorly relative to the GTR + SSR$_4$ model; yet, this model indicated relatively strong bootstrap support for the correct node (71%). The poor performance of the SSR model was also observed for both the primate and rodent mtDNA data sets. Importantly, in those data sets where the SSR models performed poorly, they tended to be no worse than the equal rates models (nevertheless, see node B in the rodent mtDNA tree).

The failure of model selection to identify the model with the highest predictive accuracy is of great concern. We suspect that the good fit of the SSR model results from the fact that this model is derived from patterns of variation observed in the data (i.e., the nonrandom distribution of variable sites among codon positions). However, the SSR models perform poorly because they ignore biological information that is essential for accurate phylogenetic reconstruction, namely, the fine-grained rate heterogeneity within rates classes (Buckley, Simon, and Chambers 2001). Although model selection was in some cases misleading, when the nested gamma and invariable sites models were considered in isolation, model selection performed well. For the gamma and invariable sites models, both the AIC and likelihood ratio tests led to selection of the same model in almost all cases (data not shown). The substitution model used for phylogenetic analysis should always be justified statistically and biologically (e.g., Burnham and Anderson 1998; Posada and Crandall 1998).

In agreement with a previous study (Buckley, Simon, and Chambers 2001), we observed that the SSR models consistently gave lower estimates of branch lengths relative to the GTR + Γ and GTR + I + Γ models (table 3), probably because of the false assumption of equal rates among sites within rate categories. This underestimation of the number of substitutions may make analyses using the SSR models more susceptible to systematic error, especially when coupled with rate heterogeneity among lineages, and we believe that this is the phenomenon that we are observing with the bird mtDNA sequences.

Invariable Sites Models Versus Mixed Gamma and Invariable Sites Models

For the bird mtDNA, the GTR + I model was the worst fitting of the among-site rate variation models, and indeed this model gave low estimates of bootstrap support for the correct root of the tree (36%). The poor performance of the GTR + I model was also observed in the primate mtDNA data set, especially the support for the chimpanzee-human clade. We also observed that the GTR + I model gave consistently lower estimates of tree length for all data sets relative to the models that incorporated a gamma distribution. The presence of rapidly evolving sites in the bird and primate data sets may be causing the invariable sites model to underestimate the rate of change over the tree, leading to selection of an incorrect topology.

Because likelihood calculations under an invariable sites model are much faster than under a gamma rates model, the use of an invariable sites model may be desirable for large data sets. Cunningham, Zhu, and Hillis (1998) argued that invariable sites models would in some cases be an acceptable approximation to the distribution of among-site rate variation even if a gamma rates model fits the data better. This seems to be the case, for example, with the arthropod EF1α sequences. However, for the bird and primate mtDNA data sets, the invariable sites model was a poor substitute for the gamma rates model (see also Buckley, Simon, and Chambers 2001), as was suggested by the AIC. In general, it seems difficult to predict when a suboptimal among-site rate variation model will be an acceptable approximation, although for the bird and primate mtDNA data sets, the AIC was a good guide if the SSR models were excluded.

When one or more substitution models have small AIC differences (e.g., $\Delta_i \leq 2.0$), it may be appropriate to perform a sensitivity analysis because the statistical justification for preferring one model to another is weak (Chatfield 1995). Because the process of model selection is essentially an estimation procedure, there will often be an uncertainty associated with selecting the best-fit model, as we have shown for several of the data sets that we have analyzed here. A good example of model selection uncertainty (Madigan and Raftery 1994; Chatfield 1995; Gelman et al. 1995; Burnham and Anderson 1998) was observed in the insect 18S rRNA data set where the AIC difference for the GTR + I + Γ model was only 2.0, indicating that the data are essentially un-

able to differentiate between the GTR + Γ and GTR + I + Γ models. Although the estimates of bootstrap support and tree length between the two models are similar, this will not always be the case (e.g., Buckley, Simon, and Chambers 2001). Although sensitivity analyses will increase the computational burden of ML phylogenetic reconstruction, we do not believe that the problem is as severe as was asserted by Sanderson and Kim (2000) because through prudent application of the AIC and biological knowledge, we can focus on a small set of reasonable models. We also note that it is very difficult to accommodate model selection uncertainty if likelihood ratio tests are used (Burnham and Anderson 1998).

Conclusions

We have presented analyses from a number of data sets that contradict the argument that complex models are of little utility in molecular phylogenetics (e.g., Takahashi and Nei 2000). For a number of known phylogenies, we have explored the complexity of the bias versus variance relationship using nonparametric bootstrapping. We observed that for these data sets, although the SSR models always had a superior fit to the data relative to invariable sites and gamma rates models, the SSR models often gave misleading estimates of phylogenetic relationships. Additionally, bootstrap support values for most of the data sets increased under more realistic modeling of the nucleotide substitution process. Although our results clearly illustrate the utility of realistic models, we still have much to learn concerning the effects of model misspecification on phylogenetic inference.

Acknowledgments

LITERATURE CITED

AKAIKE, H. 1973. Information theory as an extension of the maximum likelihood principle. Pp. 267–281 in B. N. PETROV and F. CSAKI, eds. Second international symposium on information theory. Akademiai Kiado, Budapest.

ALLARD, M. W., and M. M. MIYAMOTO. 1992. Testing phylogenetic approaches with empirical data, as illustrated with the parsimony method. Mol. Biol. Evol. **9**:778–786.

ANDRIEU, G., G. CARAUX, and O. GASCUEL. 1997. Confidence intervals of evolutionary distances between sequences and comparison with usual approaches including the bootstrap method. Mol. Biol. Evol. **14**:875–882.

BOLLBACK, J. P., and J. P. HUELSENBECK. 2001. Phylogeny, genome evolution, and host specificity of single-stranded RNA bacteriophage (Family Leviviridae). J. Mol. Evol. **52**:117–128.

BOORE, J. L., D. V. LAVROV, and W. M. BROWN. 1998. Gene translocation links insects and crustaceans. Nature **392**:667–668.

BRUNO, W. J., and A. L. HALPERN. 1999. Topological bias and inconsistency of maximum likelihood using wrong models. Mol. Biol. Evol. **16**:564–566.

BUCKLEY, T. R., C. SIMON, and G. K. CHAMBERS. 2001. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths and bootstrap support. Syst. Biol. **50**:67–86.

BURNHAM, K. P., and D. R. ANDERSON. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York.

CHATFIELD, C. 1995. Model uncertainty, data mining and statistical inference. J. R. Stat. Soc. A **158**:419–466.

CRACRAFT, J. 2001. Avian evolution, Gondwana biogeography and the cretaceous-tertiary mass extinction event. Proc. R. Soc. Lond. B **268**:459–469.

CUNNINGHAM, C. W. 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. Syst. Biol. **46**:464–478.

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.

EFRON, B. 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. **7**:1–26.

EFRON, B., E. HALLORAN, and S. HOLMES. 1996. Bootstrap confidence levels for phylogenetic trees. Proc. Natl. Acad. Sci. USA **93**:7085–7090.

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

———. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

FELSENSTEIN, J., and H. KISHINO. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. Syst. Biol. **42**:193–200.

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

FORSTER, M. R. 2000. Key concepts in model selection: performance and generalizability. J. Math. Psychol. **44**:205–231.

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.

FRIEDRICH, M., and D. TAUTZ. 1995. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. Nature **376**:165–166.

———. 2000. Arthropod rDNA phylogeny revisited: a consistency analysis using Monte Carlo simulation. Ann. Soc. Entomol. Fr. (N.S.) **37**:21–40.

GARCÍA-MONCHADO, E., M. PEMPERA, N. DENNEBOUY, M. OLIVA-SUAREZ, J.-C. MOUNOLOU, and M. MONNEROT. 1999. Mitochondrial genes collectively suggest the paraphyly of crustaceans with respect to Insecta. J. Mol. Evol. **49**:142–149.

GARCIA-MORENO, J., and D. P. MINDELL. 2000. Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): a case study using avian CHD. Mol. Biol. Evol. **17**:1826–1832.

GELMAN, A., J. B. CARLIN, H. S. STERN, and D. B. RUBIN. 1995. Bayesian data analysis. Chapman and Hall/CRC, London.

GIRIBET, G., G. D. EDGECOMBE, and W. C. WHEELER. 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature **413**:157–161.

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

GOODMAN, M., C. A. PORTER, J. CZELUSNIAK, S. L. PAGE, H. SCHNEIDER, J. SHOSHANI, G. GUNNELL, and G. P. GROVES. 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. **9**:585–598.

GROTH, J. G., and G. F. BARROWCLOUGH. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol. Phylogenet. Evol. **12**:115–123.

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.

HAYASAKA, K., T. GOJOBORI, and S. HORAI. 1988. Molecular phylogeny and evolution of primate mitochondrial DNA. Mol. Biol. Evol. **5**:626–644.

HILLIS, D. M., and J. J. BULL. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. **42**:182–192.

HUELSENBECK, J. P. 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? Syst. Biol. **47**:519–537.

HUELSENBECK, J. P., and D. M. HILLIS. 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. **42**:247–264.

HWANG, U. W., M. FRIEDRICH, D. TAUTZ, C. J. PARK, and W. KIM. 2001. Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature **413**:154–157.

INAGAKI, Y., J. B. DACKS, W. F. DOLITTLE, K. I. WATANABE, and T. OHAMA. 2000. Evolutionary relationships between dinoflagellates bearing obligate diatom endosymbionts: insight into tertiary endosymbiosis. Int. J. Syst. Evol. Microbiol. **50**:2075–2081.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–123 in H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KRAJEWSKI, C., M. G. FAIN, L. BUCKLEY, and D. G. KING. 1999. Dynamically heterogenous partitions and phylogenetic inference: an evaluation of analytical strategies with cytochrome b and ND6 gene sequences in cranes. Mol. Phylogenet. Evol. **13**:302–313.

KRISTENSEN, N. P. 1995. Forty years insect phylogenetic systematics. Hennigs Kritische Bemerkungen, and subsequent development. Zoologische Beitrage **36**:83–124.

LEE, M. S. Y. 2000. Tree robustness and clade significance. Syst. Biol. **49**:829–836.

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.

LOVETTE, I. J., and E. BERMINGHAM. 2000. C-mos variation in songbirds: molecular evolution, phylogenetic implications, and comparisons with mitochondrial differentiation. Mol. Biol. Evol. **17**:1569–1577.

MADIGAN, D., and A. E. RAFERTY. 1994. Model selection and accounting for model uncertainty in graphical models using Occam's window. J. Am. Stat. Assoc. **89**:1535–1546.

MANLY, B. F. J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. Chapman and Hall, London.

MINDELL, D. P., M. D. SORENSON, D. E. DIMCHEFF, M. HASEGAWA, J. C. AST, and T. YURI. 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst. Biol. **48**:138–152.

MURPHY, W. J., E. EIZIRILK, W. E. JOHNSON, Y. P. ZHANG, O. A. RYDER, and S. J. O'BRIEN. 2001. Molecular phylogenetics and the origins of placental mammals. Nature **409**:614–618.

NAYLOR, G., and W. BROWN. 1998. *Amphioxus* mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst. Biol. **47**:61–76.

PAGE, S. L., and M. GOODMAN. 2001. Catarrhine phylogeny: noncoding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. Mol. Phylogenet. Evol. **18**:14–25.

POSADA, D., and K. A. CRANDALL. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14**:817–818.

PRAGER, E. M., and A. C. WILSON. 1980. Phylogenetic relationships and rates of evolution in birds. Pp. 1209–1214 in R. NÖHRING, ed. Proceedings of the XVIIth International Ornithological Congress, Berlin.

REGIER, J. C., and J. W. SCHULTZ. 1997. Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. Mol. Biol. Evol. **14**:902–913.

———. 2001. Elongation factor-2: a useful gene for arthropod phylogenetics. Mol. Phylogenet. Evol. **20**:136–148.

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

———. 2000. Long-branch attraction phenomenon and the impact of among-site rate variation on rodent phylogeny. Gene **259**:177–187.

RUSSO, C. A. M., N. TAKEZAKI, and M. NEI. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol. Biol. Evol. **13**:525–536.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SANDERSON, M. J. 1989. Confidence limits on phylogenies: the bootstrap revisited. Cladistics **5**:113–129.

———. 1995. Objections to bootstrapping phylogenies: a critique. Syst. Biol. **44**:299–320.

SANDERSON, M. J., and J. KIM. 2000. Parametric phylogenetics? Syst. Biol. **49**:817–829.

SATTA, Y., J. KLEIN, and N. TAKAHATA. 2000. DNA archives and our nearest relative: the trichotomy problem revisited. Mol. Phylogenet. Evol. **14**:259–275.

SCHULTZ, J. W., and J. C. REGIER. 2000. Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean+hexapod clade. Proc. R. Soc. Lond. B **267**:1011–1019.

SHIBATA, R. 1989. Statistical aspects of model selection. Pp. 215–240 in J. C. WILLEMS, ed. From data to model. Hong Kong Baptist University, Hong Kong.

SHOSHANI, J., C. P. GROVES, E. L. SIMONS, and G. F. GUNNELL. 1996. Primate phylogeny: morphological vs molecular results. Mol. Phylogenet. Evol. **5**:102–154.

SIBLEY, C. G., and J. E. ALQUIST. 1990. Phylogeny and classification of birds: a study in molecular evolution. Yale University Press, New Haven, Conn.

SIDDALL, M. E. 1998. Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris zone. Cladistics **14**:209–220.

SIDDALL, M. E., and M. F. WHITING. 1999. Long-branch abstractions. Cladistics **15**:9–24.

STANGER-HALL, K., and C. W. CUNNINGHAM. 1998. Support for a monophyletic Lemuriformes: overcoming incongruence between data partitions. Mol. Biol. Evol. **15**:1572–1577.

STAPEL, S. O., J. A. M. LEUNISSEN, M. VERSTEEG, J. WATTEL, and W. W. DE JONG. 1984. Ratites as oldest offshoot of avian stem—evidence from α-crystallin A sequences. Nature **311**:257–259.

STEEL, M., and D. PENNY. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. Mol. Biol. Evol. **17**:839–850.

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., J. A. MARKERT, and C. W. 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.

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

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 *in* D. M. HILLIS, C. MOTITZ, and B. K. MABLE, eds. Molecular systematics. 2nd edition. Sinauer, Sunderland, Mass.

SWOFFORD, D. L., P. J. WADDELL, J. P. HUELSENBECK, P. G. FOSTER, P. O. LEWIS, and J. S. ROGERS. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood. Syst. Biol. **50**:525–539.

TAKAHASHI, K., and M. NEI. 2000. Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. Mol. Biol. Evol. **17**:1251–1258.

TOURASSE, N. J., and M. GOUY. 1999. Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. Mol. Phylogenet. Evol. **13**:159–168.

VAN TUINEN, M., C. G. SIBLEY, and S. B. HEDGES. 2000. The early history of modern birds inferred from DNA sequences of nuclear and ribosomal genes. Mol. Biol. Evol. **17**:451–457.

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., 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.

WASSERMAN, L. 2000. Bayesian model selection and model averaging. J. Math. Psychol. **44**:92–107.

WHITING, M. F., J. C. CARPENTER, Q. D. WHEELER, and W. C. WHEELER. 1997. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal sequences and morphology. Syst. Biol. **46**:1–68.

WILSON, K., V. CAHILL, E. BALLMENT, and J. BENZIE. 2000. The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are Malacostracan crustaceans more closely related to insects than to branchiopods? Mol. Biol. Evol. **17**:863–874.

XIA, X. 2000. Phylogenetic relationships among horseshoe crabs species: effect of substitution models on phylogenetic analyses. Syst. Biol. **49**:87–100.

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

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

———. 1996. Phylogenetic analysis using parsimony and likelihood methods. J. Mol. Evol. **42**:294–307.

———. 1997*a*. PAML: a program for phylogenetic analysis by maximum likelihood. CABIOS **13**:555–556.

———. 1997*b*. How often do wrong evolutionary models produce better phylogenies. Mol. Biol. Evol. **14**:105–108.

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.

ZIETKIEWICZ, E., C. RICHER, and D. LABUDA. 1999. Phylogenetic affinities of tarsier in the context of primate Alu repeats. Mol. Phylogenet. Evol. **11**:77–83.