

# Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics

Yoshiyuki Suzuki<sup>†</sup>, Galina V. Glazko, and Masatoshi Nei<sup>‡</sup>

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, 328 Mueller Laboratory, University Park, PA 16802

Contributed by Masatoshi Nei, October 23, 2002

**Bayesian phylogenetics has recently been proposed as a powerful method for inferring molecular phylogenies, and it has been reported that the mammalian and some plant phylogenies were resolved by using this method. The statistical confidence of interior branches as judged by posterior probabilities in Bayesian analysis is generally higher than that as judged by bootstrap probabilities in maximum likelihood analysis, and this difference has been interpreted as an indication that bootstrap support may be too conservative. However, it is possible that the posterior probabilities are too high or too liberal instead. Here, we show by computer simulation that posterior probabilities in Bayesian analysis can be excessively liberal when concatenated gene sequences are used, whereas bootstrap probabilities in neighbor-joining and maximum likelihood analyses are generally slightly conservative. These results indicate that bootstrap probabilities are more suitable for assessing the reliability of phylogenetic trees than posterior probabilities and that the mammalian and plant phylogenies may not have been fully resolved.**

**B**ayesian inference by using the Markov chain Monte Carlo method has been advocated as a powerful tool for inferring phylogenetic relationships of different species in the post-genomic era (1), in which it would become a common practice to construct phylogenetic trees by using concatenated sequences of a large number of genes (2). In fact, when Murphy *et al.* (3) constructed a phylogenetic tree of major mammalian species using concatenated nucleotide sequences of 22 genes, the reliability of interior branches (or clades) as judged by the posterior probability in Bayesian analysis was generally higher than that as judged by the bootstrap probability (4) in maximum likelihood (ML) analysis (5). Eleven of 27 interior branches were supported with 95% confidence level only by the posterior probability, whereas no interior branch was supported by the bootstrap probability alone. Similarly, when Karol *et al.* (6) constructed a phylogenetic tree of some major evolutionary lineages of plants using concatenated nucleotide sequences of four genes, 12 of 37 interior branches were supported only by the Bayesian posterior probability, whereas no interior branch was supported by the ML bootstrap probability alone. Murphy *et al.* (3) interpreted these results as an indication that “bootstrap support may be too conservative,” and the mammalian and plant phylogenies were claimed to have been resolved. However, it is possible that the posterior probability was too high or too liberal. The purpose of this paper is to examine which of these interpretations is more reasonable by conducting computer simulation.

Theoretically, a phylogenetic tree of genes from different species should be bifurcating, because the replication of nucleotide sequence is a bifurcating process. Therefore, if we construct a gene tree for four species *A*, *B*, *C*, and *D*, one of three possible topologies,  $((A, B), (C, D))$ ,  $((A, D), (B, C))$ , and  $((A, C), (B, D))$ , will be chosen. In reality, however, different genes from the same set of species may show different topologies because of the polymorphism, recombination, and homoplasy in ancestral populations (7). If we concatenate sequences of many genes of the same size or similar sizes and construct a phylogenetic tree, the inferred tree is likely to have the topology that is supported by the largest number of genes in the sequences. It is therefore

important to use a large number of randomly chosen genes in the inference of species phylogenies.

In the statistical inference of phylogenetic trees of four species, the null hypothesis to be tested is that the three different topologies occur with equal frequency. If a particular topology is chosen with high statistical confidence, we assume that this topology is established, although it may be rejected later by some additional data. If different species diverged during a short period of evolutionary time, as in the case of divergence of mammalian orders, it would be difficult to identify the true tree unless we use a large number of genes. However, even if we use a large number of genes without any bias in G-C content, a wrong tree may be identified as though it were the true tree (false positive) if an excessively liberal statistical method is used. Here, we examine the frequency of occurrence of this false-positive result in the Bayesian, neighbor-joining (NJ; ref. 8), and ML methods under the condition that all of the three topologies occur with equal frequency. We then discuss statistical properties of Bayesian posterior and bootstrap probabilities.

## Methods

Three sets of four nucleotide sequences ( $a', b', c',$  and  $d'$ ), ( $a'', b'', c'',$  and  $d''$ ), and ( $a''', b''', c''',$  and  $d'''$ ) with 5,000 sites were generated following topologies  $((a', b'), (c', d'))$ , (Fig. 1A);  $((a'', d''), (b'', c''))$ , (Fig. 1B); and  $((a''', c'''), (b''', d'''))$ , (Fig. 1C), respectively, by using the computer program SEGEN (version 1.25; ref. 9). The length of all exterior branches ( $b_E$ ) was assumed to be 0.05 substitutions per site and that of interior branch ( $b_I$ ) 0.005, except for a few cases in which  $b_E = 0.1$  and  $b_I = 0.01$  were assumed. These branch lengths were determined on the basis of the observation that, in a phylogenetic analysis of amino acid sequences from humans, cows, and rodents with chicken as an outgroup, different genes supported different topologies but  $b_E$  for mammalian lineages was always about 0.05 on average and  $b_I$  was about 0.005 (G.V.G. and M.N., unpublished data). In addition, the rate of nucleotide substitution seems to be similar to or slightly higher than that of amino acid substitution in mammals (10). After generating the sequences with a given model of nucleotide substitution, sequences  $a', a'',$  and  $a'''$  were concatenated into a single sequence  $a$ . Similarly,  $b', b'',$  and  $b'''$ ;  $c', c'',$  and  $c'''$ ; and  $d', d'',$  and  $d'''$  were concatenated into single sequences  $b, c,$  and  $d$ , respectively. The number of nucleotide sites in the concatenated sequences was 15,000, which was close to that (16,397) used in the phylogenetic analysis of mammalian species by Murphy *et al.* (3). Note that sequences  $a, b, c,$  and  $d$  are expected to generate three different topologies  $((a, b), (c, d))$ , (Fig. 1A);  $((a, d), (b, c))$ , (Fig. 1B); and  $((a, c), (b, d))$ , (Fig. 1C) with equal probability, but, in actual inference, one of them is chosen because of the stochastic error of nucleotide substitution. However, the inferred tree should not be supported with a high posterior or bootstrap probability, because it was

Abbreviations: NJ, neighbor-joining; ML, maximum likelihood; JC, Jukes-Cantor.

<sup>†</sup>Present address: Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, 1111 Yata, Mishima-shi, Shizuoka-ken 411-8540, Japan.

<sup>‡</sup>To whom correspondence should be addressed. E-mail: nxm2@psu.edu.

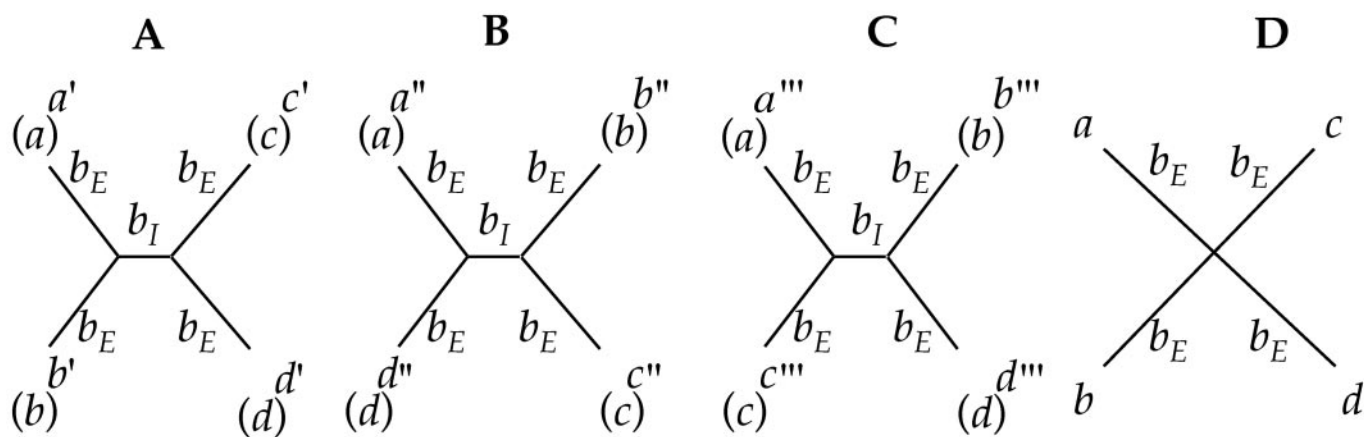


Fig. 1. Model trees used for generating concatenated sequences (A–C) and completely linked sequences (D).  $b_E$  and  $b_I$  indicate the lengths of exterior and interior branches, respectively. No interior branch exists in D.

chosen just by chance. Therefore, if it happened to be supported, the result was judged as a false-positive.

In this paper, we used simple models of nucleotide substitution to reveal the major features of Bayesian and bootstrap probabilities. The models used were the Jukes–Cantor (JC) and Kimura models, with or without rate variation among sites (11).

The phylogenetic tree of sequences  $a$ ,  $b$ ,  $c$ , and  $d$  was inferred by the Bayesian, NJ, and ML methods. We used the computer program MRBAYES (version 2.01; ref. 12) for constructing Bayesian trees. One cold and three incrementally heated chains were run for 2,000,000 generations, with random starting trees and the temperature parameter value of 0.2. Trees were sampled every 100 generations from the last 1,000,000 generations (well after the chain reached stationarity), and 10,000 sampled trees were used for inferring a Bayesian tree. MEGA (version 2.1; ref. 13) and PAUP\* (version 4.0b8a; ref. 14) were used for constructing NJ and ML trees, respectively. In both methods, we conducted 1,000 bootstrap resamplings. In the case of ML trees, the nucleotide frequencies were estimated by the observed frequencies (the

default option of PAUP\*). Bayesian, NJ, and ML trees were judged as false-positives when the posterior or bootstrap probability was  $>95\%$ . The entire procedure was repeated 50 times (replications) for estimating the false-positive rate of these methods. Note that the expected false-positive rate (type-I error) is 5% for all methods because the confidence level is 95%.

## Results

We first analyzed the concatenated sequences generated with the JC model and the expected branch lengths of  $b_E = 0.05$  and  $b_I = 0.005$  in each of the three model trees. Bayesian analysis produced topologies  $((a, b), (c, d))$ ,  $((a, d), (b, c))$ , and  $((a, c), (b, d))$  in 16 (32%), 14 (28%), and 20 (40%) replications, respectively (Table 1). These frequencies were more or less equal to one another, because the topologies were chosen with the same probability in each replication. However, the posterior probability was 85% on average and  $>95\%$  in 21 replications (42%). Therefore, the false-positive rate was much higher than the expected (5%), indicating that the Bayesian

Table 1. Frequencies of false-positives inferred by Bayesian posterior probability and NJ and ML bootstrap probabilities for JC and Kimura models

$b_E$	$b_I$	$R$	Bayesian posterior probability					NJ bootstrap probability					ML bootstrap probability				
			A	B	C	All	$\bar{P}$	A	B	C	All	$\bar{P}$	A	B	C	All	$\bar{P}$
0.05	0.005	0.5 (0.5)	8	5	8	21	85	1	0	1	2	63	0	0	0	0	64
			16	14	20	50	17	13	20	50	16	14	20	50			
0.1	0.01	0.5 (0.5)	6	7	7	20	85	1	0	0	1	64	0	0	0	0	65
			18	20	12	50	16	19	15	50	18	20	12	50			
0.05	0.005	5 (0.5)	14	13	9	36	91	0	0	0	0	62	0	0	0	0	63
			19	19	12	50	17	20	13	50	19	19	12	50			
0.1	0.01	5 (0.5)	12	14	11	37	95	0	1	1	2	68	0	0	1	1	69
			18	16	16	50	19	18	13	50	18	16	16	50			
0.05	0	5 (0.5)	11	14	6	31	89	1	0	1	2	68	1	0	1	2	66
			18	21	11	50	18	22	10	50	18	21	11	50			
0.1	0	5 (0.5)	11	13	15	39	95	0	0	1	1	69	0	0	1	1	68
			18	16	16	50	16	18	16	50	18	16	16	50			

The number of false-positive cases (replications) is given above the line, and the number of replications that supported tree A, B, or C (see Fig. 1) is given below the line. All, results for all replications;  $\bar{P}$ , average probability for all replications;  $R$ , transition/transversion ratio used for generating sequence data ( $R$  value used for phylogenetic inference is given in parentheses).

posterior probability is excessively high as an indication of statistical confidence.

By contrast, the NJ bootstrap probability was 63% on average and >95% in two replications (4%; Table 1). The average ML bootstrap probability was 64%, with no false-positives. These false-positive rates were similar to or lower than the expected rate, indicating that the NJ and ML bootstrap probabilities are slightly conservative.

In Fig. 2, the Bayesian posterior probability and the NJ and ML bootstrap probabilities obtained for the same set of sequences are plotted as a scattergram to show the relationship between them. NJ and ML bootstrap probabilities are located roughly on the diagonal line (Fig. 2C), indicating that they are similar to each other. However, the posterior probability is much higher than the bootstrap probability; the former is  $\approx 100\%$  when the latter is 70% or higher (Fig. 2A and B). Similar results were obtained when we analyzed the concatenated sequences generated under the assumption of  $b_E = 0.1$  and  $b_I = 0.01$  in the model trees (Table 1; Fig. 2D–F).

In the above analysis, we used the same (JC) model for generating and analyzing the sequences. In actual data analysis, however, we usually do not know the correct model by which the sequences were generated but assume a simplified model for analyzing them. To examine the effect of using a simplified model on the false-positive rate, we generated the sequences following Kimura's model (15), with a transition/transversion ratio ( $R$ ) of 5 and analyzed them using the JC model. Note that, in the JC model,  $R = 0.5$ . When we used  $b_E = 0.05$  and  $b_I = 0.005$  in the model trees, the Bayesian posterior probability was 91% on average and >95% in 36 replications (72%), whereas the NJ and ML bootstrap probabilities were 62% and 63% on average, respectively, both with no false-positives (Table 1). In the scattergram, two bootstrap probabilities seem to be similar to each other (Fig. 2I), but the posterior probability is  $\approx 100\%$  when the bootstrap probability is  $\approx 60\%$  or higher (Fig. 2G and H). We obtained similar results under the assumption of  $b_E = 0.1$  and  $b_I = 0.01$  in the model trees (Table 1; Fig. 2J–L). These results indicate that the posterior probability is unreasonably high in the analysis of concatenated sequences whereas the bootstrap probability is still slightly conservative.

We then generated unconcatenated nucleotide sequences with 15,000 sites following the star phylogeny (Fig. 1D) using the Kimura model with  $R = 5$  and analyzed them using the JC model, to examine the effect of using a simplified model on the phylogenetic analysis of completely linked sequences. Note that the star phylogeny is usually used as the null hypothesis tree in statistical inference of phylogenetic trees of completely linked sequences (11). When we used  $b_E = 0.05$  and  $b_I = 0$  in the model trees, the Bayesian posterior probability was 89% on average and >95% in 31 replications (62%; Table 1). By contrast, the NJ and ML bootstrap probabilities were 68% and 66% on average, respectively, and >95% in two replications (4%). In the scattergram, NJ and ML bootstrap probabilities are again similar to each other (Fig. 2O), whereas the posterior probability is  $\approx 100\%$  when the bootstrap probability is 70% or higher (Fig. 2M and N). Similar results were also obtained when we used  $b_E = 0.1$  and  $b_I = 0$  (Table 1; Fig. 2P–R), indicating that the posterior probability is excessively high even in the analysis of completely linked sequences whereas the bootstrap probability is again slightly conservative.

In actual DNA sequences the evolutionary rate varies from nucleotide site to nucleotide site, and this variation is usually approximated by a gamma ( $\Gamma$ ) distribution (16). We therefore conducted another simulation using the JC +  $\Gamma$  and the Kimura +  $\Gamma$  model. First, we generated concatenated sequences using the JC +  $\Gamma$  model with a gamma parameter value ( $a$ ) of 1 and the expected branch lengths of  $b_E = 0.05$  and  $b_I = 0.005$  and inferred Bayesian, NJ, and ML trees using the

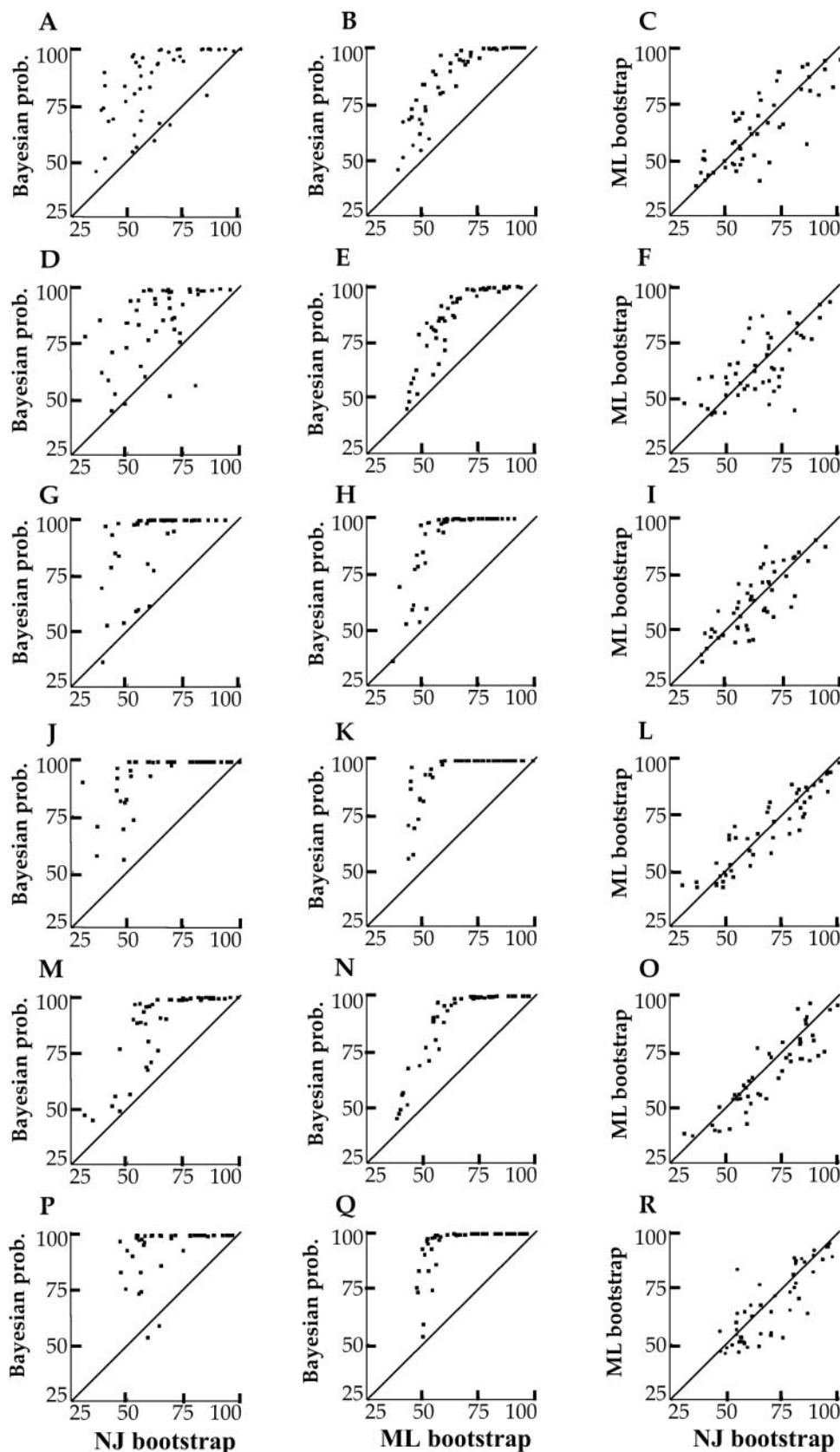
same JC +  $\Gamma$  model. In this case, topologies  $A$ ,  $B$ , and  $C$  were obtained again with nearly the same frequencies in all Bayesian, NJ, and ML analyses (data not shown). In the case of Bayesian analysis, however, the posterior probability was >95% in 22 of the 50 replications (44%), and the average probability value for all replications ( $\bar{P}$ ) was 84% (Table 2). By contrast, the bootstrap probability was >95% only in two replications in NJ analysis and only in one replication in ML.  $\bar{P}$  was 63% in NJ and 65% in ML. When we generated sequence data using the Kimura +  $\Gamma$  model with  $R = 5$  and  $a = 1$  and constructed Bayesian, NJ, and ML trees using the same model, the results were nearly the same (Table 2). In NJ, the false positive rate (10%) was higher than that (4%) for the case of the JC +  $\Gamma$  model probably by chance, but the  $\bar{P}$  value (66%) was nearly the same as that for the latter case or the cases considered in Table 1.

When the sequence data were generated by the JC +  $\Gamma$  model ( $R = 0.5$  and  $a = 1$ ) but phylogenetic inference was done with the JC model ( $R = 0.5$  and  $a = \infty$ ), the false positive rate was 41/50 or 82% and  $\bar{P}$  was 95% for Bayesian analysis. By contrast, the false positive rate was 4% and  $\bar{P}$  was 64–65% for NJ and ML. When the sequences were generated by the Kimura +  $\Gamma$  model ( $R = 5$  and  $a = 1$ ) and trees were inferred with either the JC ( $R = 0.5$  and  $a = \infty$ ) or the Kimura ( $R = 5$  and  $a = \infty$ ) model, the false positive rate and the  $\bar{P}$  value were essentially the same as those for the above case. Therefore, the false-positive rate is too high in Bayesian analysis, whereas it is close to the expected value (5%) in NJ and ML. Tables 1 and 2 show that both NJ and ML bootstrap tests tend to be slightly conservative but that the NJ test is not always so.

## Discussion

We demonstrated that the posterior probability in Bayesian phylogenetics can be excessively high in the analysis of concatenated sequences even when the same model as that for generating each gene sequence was used. The false-positive rate became even higher when a simplified model was used for phylogenetic inference. Under the same condition, the posterior probability was also excessively liberal in the analysis of completely linked sequences. In actual data analysis, we usually do not know the correct model by which the sequences were generated but use a simplified model for analyzing them, as mentioned above. The posterior probability therefore can be unreasonably high in actual data analysis even when unconcatenated sequences are used.

By contrast, the bootstrap probabilities in NJ and ML analyses were generally slightly conservative regardless of whether the correct or simplified model was used or whether the concatenated or completely linked sequences were analyzed. This is particularly so for ML analysis. The bootstrap probability therefore seems to be a conservative estimate of statistical confidence. These results are consistent with the previous observations that the false-positive rate of bootstrap probabilities in the NJ and maximum parsimony methods is lower than the expected rate in phylogenetic analysis of unconcatenated sequences, as long as these methods are not inconsistent (17–20). In fact, the bootstrap probability, when it is close to unity, is theoretically shown to be an underestimate if it is simply interpreted as the probability that the inferred tree is correct (21). However, a conservative method should be preferable to an overly liberal method in phylogenetic analysis, because we usually draw conclusions only from statistical analysis without doing any experiments (11). In addition, it may be possible to modify the resampling procedure for obtaining a less conservative value of bootstrap probability (22). The bootstrap probability therefore seems to be more suitable for assessing the reliability of phylogenetic



**Fig. 2.** Scattergrams of Bayesian posterior probability and NJ bootstrap probability (A, D, G, J, M, and P), posterior probability and ML bootstrap probability (B, E, H, K, N, and Q), and two bootstrap probabilities (C, F, I, L, O, and R) obtained for the same set of sequences. The sequences analyzed are (i) concatenated sequences generated by using the model trees (Fig. 1 A–C) with  $b_E = 0.05$ ,  $b_I = 0.005$  and  $R = 0.5$  (A–C);  $b_E = 0.1$ ,  $b_I = 0.01$ , and  $R = 0.5$  (D–F);  $b_E = 0.05$ ,  $b_I = 0.005$  and  $R = 5$  (G–I); and  $b_E = 0.1$ ,  $b_I = 0.01$ , and  $R = 5$  (J–L) and (ii) completely linked sequences generated by using the model tree (Fig. 1D) with  $b_E = 0.05$  (M–O) and  $b_E = 0.1$  (P–R). Diagonal lines indicate the perfect correspondence between two probabilities.

**Table 2. Frequencies of false-positives inferred by Bayesian posterior probability and NJ and ML bootstrap probabilities for JC +  $\Gamma$  and Kimura +  $\Gamma$  models**

$b_E$	$b_I$	$R$	$a$	Bayes. prob.		NJ boot. prob.		ML boot. prob.	
				All reps.	$\bar{P}$	All reps.	$\bar{P}$	All reps.	$\bar{P}$
0.05	0.005	0.5 (0.5)	1 (1)	22/50	84	2/50	63	1/50	65
0.05	0.005	5 (5)	1 (1)	16/50	83	5/50	66	2/50	62
0.05	0.005	5 (0.5)	1 (1)	35/50	94	1/50	63	0/50	63
0.05	0.005	0.5 (0.5)	1 ( $\infty$ )	41/50	95	2/50	65	2/50	64
0.05	0.005	5 (0.5)	1 ( $\infty$ )	40/50	96	3/50	66	2/50	65
0.05	0.005	5 (5)	1 ( $\infty$ )	35/50	92	3/50	65	2/50	65

The number of false-positive cases (replications) is given before the slash sign, and the total number of replications is given after the slash sign. All reps., results from all replications; boot. prob., bootstrap probability;  $\bar{P}$ , average probability for all replications;  $R$ , transition/transversion ratio used for generating sequence data ( $R$  value used for phylogenetic inference is given in parentheses);  $a$ , gamma parameter used for generating sequence data ( $a$  value used for phylogenetic inference is given in parentheses).

trees than the posterior probability, although the theoretical basis of bootstrap probability is not well understood at present.

A high Bayesian posterior probability for a given interior branch (or a clade) is obviously due to the appearance of the same branch or clade in most or all sampled trees used for constructing a consensus tree. This result is in turn caused by the fact that the highest ML tree (or set of trees) is visited again and again in the Markov chain Monte Carlo computation for the original set of sequences. In the computation of bootstrap probability, however, the original set of sequences is considered as a single evolutionary event realized with stochastic errors, and therefore the original sequences are reshuffled (bootstrap-resampled) to evaluate the reliability of the original or consensus tree. In this case, different bootstrap-resampled sequences may generate different ML or NJ trees, unless the extent of stochastic errors is small. Therefore, the bootstrap probability computed for a bootstrap consensus tree is expected to be lower than the Bayesian posterior probability. Because original sequences are always subject to stochastic errors, the reliability of an inferred tree should be evaluated by considering stochastic errors.

In the phylogenetic analyses of mammals (3) and plants (6), some interior branches were not supported by ML bootstrap probabilities, as mentioned above. Therefore, we suspect that the phylogenetic trees published in these papers may not have been established yet. Similarly, the reliability of other molecular phylogenies obtained by Bayesian phylogenetics (e.g., refs. 23–25) should be reexamined by additional methods and additional sequence data. It is known that two different topologies for the same set of mammalian species can both be supported by high Bayesian probabilities when DNA and protein data were analyzed separately (K. Misawa and M.N., unpublished work). This finding also indicates that Bayesian phylogenetics may give overcredibility of the tree inferred.

### Addendum

After completion of this paper, we came to know a paper in which a computer simulation was conducted to evaluate the reliability of Bayesian and bootstrap probabilities as a statistical confidence of interior branches (26). The model tree used was a maximum likelihood tree of 23 species of snakes

obtained by using the general time reversible (GTR) model of nucleotide substitution with invariable sites (I) plus a  $\Gamma$  distribution of variable-rate sites (GTR + I +  $\Gamma$  model) for a portion of mitochondrial DNA. By using this model tree and the same GTR + I +  $\Gamma$  model, 120 datasets of 500 nucleotide sites were randomly generated, and each of the datasets was used to infer Bayesian and ML trees. This computation generated  $(23 - 3) \times 120 = 2400$  Bayesian posterior probability (PP) and bootstrap probability (BP) values ranging from 0% to 100%. These probability values were then classified into 10 bins with an interval of 10%. At the same time, the proportions of interior branches that were correctly inferred among 120 reconstructed trees (PBC) were computed for each bin class of PP and BP values. Comparison of PBC and PP or BP showed that BP is a clear underestimate of PBC though PP also tends to be an underestimate. From this observation, the authors concluded that PP is a better indicator of statistical confidence than BP.

This simulation is different from ours in that the same substitution model as that for sequence generation was used for phylogenetic inference without concatenation of different genes that might have evolved differently. In reality, the substitution model used for phylogenetic inference would never be the same as the real substitution pattern, and many different genes are concatenated when a large-scale data analysis is done. Therefore, we believe our simulation is more realistic than the one mentioned above. It should also be noted that the null hypothesis of the statistical tests used in the above simulation is not clearly defined, though theoretically it should be the absence (or length 0) of the interior branch under consideration (11). When a tree with many positive interior branches is used as a model tree for generating replicate datasets, the bootstrap test of this null hypothesis is quite complicated (19, 21). More theoretical studies are needed on this important problem.

We thank Dan Graur, Xun Gu, Rodney Honeycutt, Junhyong Kim, Sudhir Kumar, Bill Martin, Mike Miyamoto, Pam Soltis, and Jianzhi (George) Zhang for their comments on an earlier version of the manuscript. This work was supported by a grant from the National Institutes of Health (GM20293) to M.N.

- Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. (2001) *Science* **294**, 2310–2314.
- Nei, M., Xu, P. & Glazko, G. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 2497–2502.
- Murphy, W. J., Eizirik, E., O'Brien, S. J., Madsen, O., Scally, M., Douady, C. J., Teeling, E., Ryder, O. A., Stanhope, M. J., deJong, W. W. & Springer, M. S. (2001) *Science* **294**, 2348–2351.
- Felsenstein, J. (1985) *Evolution* **39**, 783–791.
- Felsenstein, J. (1981) *J. Mol. Evol.* **17**, 368–376.
- Karol, K. G., McCourt, R. M., Cimino, M. T. & Delwiche, C. F. (2001) *Science* **294**, 2351–2353.
- Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
- Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
- Rambaut, A. & Grassly, N. C. (1997) *Comput. Appl. Biosci.* **13**, 235–238.
- Wolfe, K. H. & Sharp, P. M. (1993) *J. Mol. Evol.* **37**, 441–456.
- Nei, M. & Kumar, S. (2000) *Molecular Evolution and Phylogenetics* (Oxford Univ. Press, Oxford).

12. Huelsenbeck, J. P. & Ronquist, F. (2001) *Bioinformatics* **17**, 754–755.
13. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001) *Bioinformatics* **17**, 1244–1245.
14. Swofford, D. L. (2000) *PAUP\*: Phylogenetic Analysis Using Parsimony (and Other Methods)* (Sinauer, Sunderland, MA).
15. Kimura, M. (1980) *J. Mol. Evol.* **16**, 111–120.
16. Jin, L. & Nei, M. (1990) *Mol. Biol. Evol.* **7**, 82–102.
17. Zharkikh, A. & Li, W.-H. (1992) *J. Mol. Evol.* **35**, 356–366.
18. Zharkikh, A. & Li, W.-H. (1992) *Mol. Biol. Evol.* **9**, 1119–1147.
19. Sitnikova, T., Rzhetsky, A. & Nei, M. (1995) *Mol. Biol. Evol.* **12**, 319–333.
20. Hillis, D. M. & Bull, J. J. (1993) *Syst. Biol.* **42**, 182–192.
21. Efron, B., Halloran, E. & Holmes, S. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 7085–7090.
22. Zharkikh, A. & Li, W.-H. (1995) *Mol. Phylogenet. Evol.* **4**, 44–63.
23. Lutzoni, F., Pagel, M. & Reeb, V. (2001) *Nature* **411**, 937–940.
24. Leache, A. D. & Reeder, T. W. (2002) *Syst. Biol.* **51**, 44–68.
25. Buckley, T. R., Arensbarger, P., Simon, C. & Chambers, G. K. (2002) *Syst. Biol.* **51**, 4–18.
26. Wilcox, T. P., Zwickl, D. J., Heath, T. A. & Hillis, D. M. (2002) *Mol. Phylogenet. Evol.* **25**, 361–371.