

Letter to the Editor

Traditional Phylogenetic Reconstruction Methods Reconstruct Shallow and Deep Evolutionary Relationships Equally Well

Michael S. Rosenberg and Sudhir Kumar

Department of Biology, Arizona State University

The wealth of data available for molecular phylogenetic analyses is expanding at an exponential pace. As data sets have become larger, it has become increasingly critical to understand the advantages and disadvantages of using various phylogenetic inference methods. Four inference methods based on three optimization criteria are commonly used to reconstruct evolutionary history from molecular data: neighbor joining (NJ), minimum evolution (ME), maximum parsimony (MP), and maximum likelihood (ML). The overall efficiency and performance of these methods in reconstructing the true tree is known to vary with substitution rate, transition-transversion ratio, and sequence divergence (Miyamoto and Cracraft 1991; Nei and Kumar 2000).

Computer simulation has proven to be an excellent means of assessing the performance of tree-building methods (reviewed in Nei and Kumar 2000, chapter 9). It can be used to examine the overall performance of a method or specific aspects of its performance (e.g., Hillis 1996; Strimmer and von Haeseler 1996; Kim 1998; Nei, Kumar, and Takahashi 1998). Kumar and Gadagkar (2000) extended the use of simulation to examine the relative performance of NJ in reconstructing deep and shallow nodes. They found that NJ reconstructs deep branches as efficiently as shallow branches, where branch depth is defined as the minimum number of terminal taxa found between the two subtrees joined by the branch. This measure is appropriate because it estimates the complexity of inferring a branch by defining the minimum number of taxa that must be joined; deeper branches require more taxa and thus have higher complexities than shallow branches (Kumar and Gadagkar 2000). In this paper, we expand that study to include the ME, MP, and ML methods to examine how these methods perform under Jukes-Cantor (JC) model (Jukes and Cantor 1969) and a more complex Hasegawa, Kishino, and Yano (1985) (HKY) model of nucleotide substitution. We also look at how differences in choices within an inference method, e.g., the use of step matrix versus unweighted parsimony or Jukes-Cantor versus Tamura-Nei distance, affects the outcome. These simulations will allow us to examine the relative efficiencies among branch depths and among phylogenetic inference methods.

Key words: phylogenetic inference, maximum parsimony, maximum likelihood, minimum evolution, neighbor joining, deep versus shallow branches.

Address for correspondence and reprints: Sudhir Kumar, Life Sciences A-371, Department of Biology, Arizona State University, Tempe, Arizona 85287-1501. E-mail: s.kumar@asu.edu.

Mol. Biol. Evol. 18(9):1823–1827. 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

We simulated DNA sequence evolution on model trees using both the Jukes-Cantor and the HKY models of nucleotide substitution. All simulations were repeated with sequence lengths of 200 and 500 sites. The overall substitution rates (measured as the number of substitutions per site) covered a 10-fold difference. For JC, we used five substitution rates, ranging from $r = 0.00625$ to $r = 0.0625$, with three intermediate rates ($r = 0.01875$, 0.03125 , and 0.046875). For HKY, we used three substitution rates, ranging from $r = 0.0035$ to $r = 0.035$, with an intermediate rate of $r = 0.0175$. We used slower substitution rates for the HKY simulations in order to prevent the largest distances from becoming undefined due to extreme divergence. The JC and HKY results are comparable because there is a large degree of overlap between their rate ranges. For the HKY simulations, transition/transversion ratios and nucleotide frequencies were chosen after Takahashi and Nei (2000). Two transition/transversion rate ratios, $R = 2$ and $R = 5$, and two sets of nucleotide frequencies were used. In the first, values of $g_A = 0.15$, $g_C = 0.35$, $g_G = 0.15$, and $g_T = 0.35$ were used; in the second, values of $g_A = 0.10$, $g_C = 0.40$, $g_G = 0.10$, and $g_T = 0.40$ were used.

Model trees for each simulation were obtained following Kumar and Gadagkar (2000). Templates from each of two six-taxon models (A and C from Kumar and Gadagkar 2000) were concatenated to construct trees containing 6, 18, 36, 54, or 96 taxa. The expected interior branches in both A and C families of trees were kept equal, allowing one to study the performance of the methods in reconstructing branches as a function of depth alone (Kumar and Gadagkar 2000). The HKY simulations were performed only on model trees of 18, 36, and 54 taxa; the tips of these trees were shortened to reduce the maximum pairwise distance between species in order to prevent the Tamura-Nei distance (Tamura and Nei 1993) from becoming undefined.

Simulations were performed by calculating the expected number of substitutions on each branch according to the substitution model, picking a random number from a Poisson distribution with mean equal to that expected to determine the number of realized substitutions (Kumar and Gadagkar 2000), and then randomly picking sites and allowing them to change according to the model. Root sequences were random, with expected base frequencies being equal for JC or matching the frequencies listed above for HKY. After each simulation, the resultant sequences were examined in a pairwise manner for extreme divergence. Under JC, if the proportion of divergent nucleotides (p -distance) between any pair was ≥ 0.75 , the Jukes-Cantor distance could not be calculated and the simulation was repeated. A similar procedure was used for the HKY simulations to allow the calcu-

lation of the Tamura-Nei distance. The resultant tip sequences were then used for analysis in PAUP* 4.0b4a for Windows (Swofford 2000).

For the JC simulations, the NJ, ME, MP, and ML methods were all used to reconstruct the phylogenies. In this case, various combinations of model tree, sequence length, substitution rate, and tree size led to 100 sets of parameters. Each set consisted of 500 replicate simulations. A single set (model C, 96 taxa, $r = 0.0625$, 200 sites) was dropped because all simulations produced at least one pair of tip sequences with p -distance ≥ 0.75 . Jukes-Cantor distances were used for NJ and ME, and a JC model was used for ML; MP was strictly unweighted.

For HKY, the various combinations led to 144 sets of parameters. In this case, the phylogenies were estimated under a variety of methods. ME was used to estimate the phylogeny using the both the Tamura-Nei distance (ME-TN) and the Jukes-Cantor distance (ME-JC). Similarly, ML was used to estimate the phylogenies using an HKY model and a JC model (ML-HKY and ML-JC, respectively). MP was used to reconstruct the phylogeny in both an unweighted fashion (MP-UW) and a weighted fashion (MP-SM). The weighted parsimony was based on a step matrix estimated using the inverse elements of the HKY transition matrix, which was used to simulate the data. All in all, eight different methods of phylogeny reconstruction were used for the HKY simulations.

For all variations on ME, MP, and ML, a single heuristic search was performed with nearest-neighbor interchange (NNI) branch swapping. For ME and ML, the NJ tree was used as the starting tree for the NNI search; for MP, a stepwise addition procedure was used to generate the initial tree. We employed fast heuristic searches because it has been shown that the true tree is rarely the most optimal tree (Nei, Kumar, and Takahashi 1998) and that extensive searching often leads to poorer phylogenetic inference (Takahashi and Nei 2000). The maximum number of trees that could be saved during the heuristic search procedures was set to 10,000 (most of the searches never came close to reaching this limit). When multiple trees were found in one search, a majority rule consensus tree (with the option to keep groups with $<50\%$ if they were compatible with the tree) was used to create a single resultant tree for each analysis; this approach was usually used only for MP, but it was used in ME and ML in rare instances. The ability of a method to reconstruct the correct tree was evaluated by the percentage of replicates in which each correct partition appeared.

In order to make comparisons among methods, a paired-comparisons t -test (Sokal and Rohlf 1995, p. 352) was performed for each simulation data set, with a pair of methods as one factor and branch depth as the other. This test is a special case of a two-way ANOVA and differs from a standard t -test in taking into account the fact that each value within one of the groups being compared is perfectly paired with a value in the other group. In our case, the average reconstruction efficiencies for branches at a specific depth are paired among

methods for a given topology. Similar comparisons were performed to test whether there were differences in reconstruction efficiency for variation in tree shape (model A vs. model C trees), number of sites (500 vs. 200), and, for the HKY simulations, pairs of transition/transversion ratios and nucleotide frequency divergences.

Although it may seem that accuracies among branch depths are not independent, with topological errors cascading through the tree, this intuitive argument has been shown to be incorrect (Kumar and Gadagkar 2000). This is largely due to the fact that ME, MP, and ML optimize globally across an entire tree; topological errors within clades do not affect higher-level clustering as long as the monophyly of the group is inferred correctly.

Here, we report the reconstruction efficiency of a method, which is the percentage of times that a method correctly reconstructs nodes at a specific branch depth. Under JC, shallow and deep branches were reconstructed with equal efficiency for all reconstruction methods (fig. 1A). There were little differences among trees of different sizes. Comparisons among the methods (fig. 1B) revealed that ML tended to reconstruct nodes across all depths correctly slightly more often than NJ, ME, and MP. The 10% of the time that NJ, ME, and MP performed better than ML represents the slowest substitution rates. There was virtually no difference in the reconstruction efficiencies of ME and NJ. In 70% of the simulations, there was no difference between MP and ME/NJ. MP was better than ME/NJ in about 20% of the cases, and ME/NJ was better than MP in the remaining 10%. Although not overwhelmingly consistent, MP tended to outperform ME and NJ in the simulations with the slowest substitution rates, and ME and NJ tended to outperform MP in the simulations with the fastest substitution rates.

The results for the HKY simulations are similar to those for the JC simulations. All of the methods except for MP-SM reconstructed shallow and deep branches equally well (fig. 2A). Only MP-SM tended to show a bias in efficiency; the percentages of correct nodes tended to be higher for shallow branches than deep branches. The overall efficiencies of MP-SM were also much lower than those for any other method. For comparison among methods, we first looked at the effect of using different substitution models within a basic method (fig. 2B). The results were striking and clear: ME-JC was always as good as or better than ME-TN, MP-UW was always better than MP-SM, and ML-HKY was always as good as or better than ML-JC. This suggests that ME and MP work better with simpler models (even when the true substitution model is more complex), but ML works better with a complex model (see also Takahashi and Nei 2000). We then examined the efficiencies among the best variants of each method (ME-JC, MP-UW, and ML-HKY). These results were similar to those for the JC simulations. ML-HKY was better than ME-JC and MP-UW about 40% of the time, and ME-JC and MP-UW were better than ML-HKY about 30% of the time. Again, ML reconstructed correct nodes more often at higher substitution rates, and MP and ME did so at

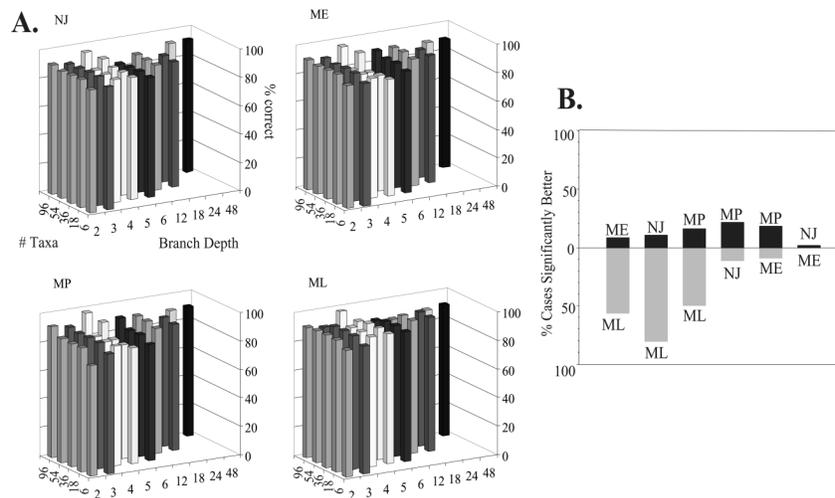


FIG. 1.—Results of JC simulations. *A*, The percentages of correct partitions as a function of branch depth for trees of different sizes. *B*, Comparison of reconstruction methods; the bars represent the percentages of conditions (parameter sets) where the reconstruction efficiency (measured as the percentage of correct partitions per branch depth) for one reconstruction method was significantly higher than that of another method.

slower substitution rates. For these analyses, ME-JC outperformed MP-UW more often than vice versa. At the intermediate rate (roughly equivalent to the second slowest rate in the JC simulations), MP-UW tended to outperform ME-JC. At the slowest rate, when there were numerous zero branch lengths in the realized trees, ME-JC outperformed MP-UW.

Under both JC and HKY, all methods were better able to reconstruct model C trees than model A trees

(results not shown). Similarly, reconstruction efficiency increased with the number of sites. Under HKY, reconstruction efficiency was consistently higher with lower transition/transversion ratios and with more uniform nucleotide frequencies.

Therefore, for the trees and parameters we studied, our results indicate that, with the exception of step-matrix weighted parsimony, none of the methods show differential ability to reconstruct shallow and deep branch-

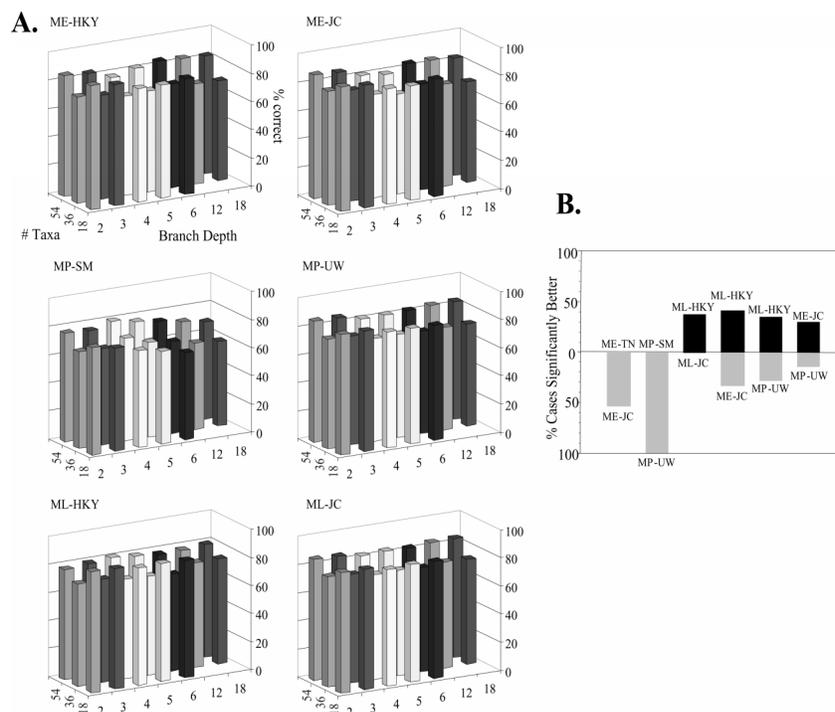


FIG. 2.—Results of HKY simulations. *A*, The percentages of correct partitions as a function of branch depth for trees of different sizes. *B*, Comparison of reconstruction methods; the bars represent the percentages of conditions (parameter sets) where the reconstruction efficiency (measured as the percentage of correct partitions per branch depth) for one reconstruction method was significantly higher than that of another method.

es. When there are differences among the methods, it is because one method is consistently better or worse at inferring phylogenetic relationships at all depths. For the trees studied, ML tends to reconstruct correct nodes more often than other methods do, particularly when the substitution rate is high. This result is similar to the results obtained by Hasegawa and Yano (1984) and Kuhner and Felsenstein (1994). MP reconstructed correct nodes more often than did the other methods at the slowest substitution rate under the JC model of nucleotide substitution, but not under the more realistic HKY model. At the faster substitution rates, MP was often the worst method. Previous studies have usually found that MP performs worse than other methods (Hasegawa and Yano 1984; Li et al. 1987; Sourdis and Nei 1988; Saitou and Imanishi 1989; Jin and Nei 1990; Kuhner and Felsenstein 1994), although Takahashi and Nei (2000) found that MP performed better than NJ at slow substitution rates. In the HKY simulations, we found that ME performed better with simple distances (JC) than with correct, complex distances (TN); a similar result was found by Takahashi and Nei (2000). ML, however, performs better with the correct model than with a simpler, incorrect model. This result is similar to the results of Yang (1996), although he also found that complex models (HKY) in ML did not work well when the true evolutionary pattern was simple (JC). Branch lengths and tree size showed little effect on the reconstruction performance of these methods. The results of our study cannot be directly compared with other work because we evaluated performance on a nodal basis rather than as a function of the entire tree (either as the percentage of correct trees or as the average distance of reconstructed trees from the correct tree). There is also a large degree of variation in various parameter choices such as tree size and shape, substitution rates, and substitution model across the studies.

The complete failure of step-matrix parsimony in our simulations was striking. In our computer simulations, this method usually reconstructed the correct nodes with less efficiency than all other methods under the corresponding conditions. It also showed the greatest sensitivity to changes in branch depth, transition/transversion ratio, and nucleotide frequency imbalance. Broughton, Stanley, and Durrett (2000) have recently examined weighted parsimony with respect to differential weighting of transitions and transversions (the most common form of step-matrix parsimony). They found that this method not only fails to discriminate phylogenetic signal from noise in most cases, but will actually add noise and make phylogenetic analyses worse in many instances. Our results agree with these results, taking them a step further since our weights included information about nucleotide frequency as well as transition/transversion ratio. We cannot overemphasize the lack of resolution provided by the use of a step-matrix in these analyses.

The simulation results given in this paper now extend the conclusions of Kumar and Gadagkar (2000) to all three major optimality criteria under both simple and complex models of nucleotide substitution and provide

a universal baseline profile for these methods. We are now examining trees with more realistic shapes and structures, variation in rates across sites, and issues such as taxon and character sampling and homogeneity of evolutionary process.

Acknowledgments

We thank Sudhindra Gadagkar and Tom Dowling and two anonymous reviewers for their comments on an earlier version of this paper. This research was supported by grants from the National Science Foundation, the National Institute of Health, and the Burroughs Wellcome Fund to S.K.

LITERATURE CITED

- BROUGHTON, R. E., S. E. STANLEY, and R. T. DURRETT. 2000. Quantification of homoplasy for nucleotide transitions and transversions and a reexamination of assumptions in weighted phylogenetic analysis. *Syst. Biol.* **49**:617–627.
- HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**:160–174.
- HASEGAWA, M., and T. YANO. 1984. Maximum likelihood method of phylogenetic inference from DNA sequence data. *Bull. Biometr. Soc. Jpn.* **5**:107.
- HILLIS, D. M. 1996. Inferring complex phylogenies. *Nature* **383**:130–131.
- JIN, L., and M. NEI. 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol. Biol. Evol.* **7**:82–102.
- JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 in H. N. MUNRO, ed. *Mammalian protein metabolism*. Academic Press, New York.
- KIM, J. 1998. Large-scale phylogenies and measuring the performance of phylogenetic estimators. *Syst. Biol.* **47**:43–60.
- 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., and S. R. GADAGKAR. 2000. Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies. *J. Mol. Evol.* **51**:544–553.
- LI, W.-H., K. H. WOLFE, J. SOURDIS, and P. M. SHARP. 1987. Reconstruction of phylogenetic trees and estimation of divergence items under non-constant rates of evolution. *Cold Spring Harb. Symp. Quant. Biol.* **52**:847–856.
- MIYAMOTO, M. M., and J. L. CRACRAFT. 1991. Phylogenetic inference, DNA sequence analysis, and the future of molecular systematics. Pp. 3–17 in M. M. MIYAMOTO and J. L. CRACRAFT, eds. *Recent advances in phylogenetic studies of DNA sequences*. Oxford University Press, Oxford, England.
- NEI, M., and S. KUMAR. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, Oxford, England.
- NEI, M., S. KUMAR, and K. TAKAHASHI. 1998. The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. *Proc. Natl. Acad. Sci. USA* **95**:12390–12397.
- SAITOU, N., and T. IMANISHI. 1989. Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.* **6**:514–525.

- SOKAL, R. R., and F. J. ROHLF. 1995. *Biometry*. 3rd Edition. W. H. Freeman and Company, New York.
- SOURDIS, J., and M. NEI. 1988. Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. *Mol. Biol. Evol.* **5**: 298–311.
- STRIMMER, K., and A. VON HAESELER. 1996. Accuracy of neighbor joining for n -taxon trees. *Syst. Biol.* **45**:516–523.
- SWOFFORD, D. L. 2000. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.
- 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.
- TAMURA, K., and M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**:512–526.
- YANG, Z. 1996. Phylogenetic analysis using parsimony and likelihood methods. *J. Mol. Evol.* **42**:294–307.

BRANDON GAUT, reviewing editor

Accepted May 4, 2001