An Investigation of Phylogenetic Likelihood Methods

Tiffani L. Williams and Bernard M.E. Moret
Department of Computer Science
University of New Mexico
Albuquerque, NM 87131-1386
Email: {tlw,moret}@cs.unm.edu

Abstract— **We analyze the performance of likelihood-based approaches used to reconstruct phylogenetic trees. Unlike other techniques such as Neighbor-Joining (NJ) and Maximum Parsimony (MP), relatively little is known regarding the behavior of algorithms founded on the principle of likelihood. We study the accuracy, speed, and likelihood scores of four representative likelihood-based methods (fastDNAml, MrBayes, PAUP*-ML, and TREE-PUZZLE) that use either Maximum Likelihood (ML) or Bayesian inference to find the optimal tree. NJ is also studied to provide a baseline comparison. Our simulation study is based on random birth-death trees, which are deviated from ultrametricity, and uses the Kimura 2-parameter + Gamma model of sequence evolution. We find that MrBayes (a Bayesian inference approach) consistently outperforms the other methods in terms of accuracy and running time.**

I. Introduction

Evolutionary biology is founded on the concept that organisms share a common origin and have subsequently diverged through time. Phylogenies (typically formulated as trees) represent our attempts to reconstruct evolutionary history. Phylogenetic analysis is used in all branches of biology with applications ranging from studies on the origin of human populations to investigations of the transmission patterns of HIV [14], and beyond, with a variety of uses in drug discovery, forensics, and security [1]. The accurate estimation of evolutionary trees is a challenging computational problem. For a given set of organisms (or taxa), the number of possible evolutionary trees is exponential. There are more than two million different trees for 10 taxa, and more than 2×10^{182} different trees for 100 taxa. An exhaustive search through the tree space is certainly not an option. Thus, scientists have designed a plethora of heuristics to assist them with phylogenetic analysis. Here, we study the performance of likelihood-based approaches for reconstructing phylogenetic trees.

Likelihood is the probability of observing the data given a particular model. It is represented mathematically as $Pr(D|M)$, read "probability of D given M". Here, D is the observed data and M is the model [18]. Different models may make the observed data more or less probable. For instance, suppose you toss a fair coin. After 100 independent tosses, you observe *heads* once and *tails* 99 times. Under the assumed model of a fair coin, the likelihood of this event is $\binom{100}{1} \cdot \frac{1}{2^{100}}$ or 7.89×10^{-29}. For a coin tossed 100 times, there are 2^{100} different, equally-likely possibilities. Only 100 of those sequences yield a single `heads`. Hence, the above event is not very likely if a fair coin is used. A more likely event is the appearance of 50 `heads` and 50 `tails`, which has a likelihood of $\binom{100}{50} \cdot \frac{1}{2^{100}}$ or 7.96×10^{-2}. Many biologists prefer likelihood approaches since they are statistically well-founded, possibly producing the most accurate phylogenetic trees. However, techniques such as Neighbor-Joining (NJ) [22] and Maximum Parsimony(MP) are also popular, since they are relatively fast, whereas likelihood-based approaches are extremely slow, limiting their use to small problems.

Many of the reconstruction methods used by biologists (distance methods, parsimony search, or quartet puzzling) have been studied extensively in an experimental environment [15], [20], [25]. Yet, very little is known concerning the behavior of likelihood-based methods. Previous work has studied the performance of such algorithms in a limited context, where typically a single criterion is of interest (i.e, number of taxa, execution time, or accuracy) [3], [7], [11], [12], [20].

Our study is the first (to the best of our knowledge) to examine the behavior of likelihood methods under a variety of different conditions. Specifically, our study varies the number of taxa (20, 40, 60), sequence length (100, 200, 400, 800, 1600), and evolutionary distance (0.1, 0.3, and 1.0). Space constraints force us to refer the reader to [27] for performance results with an evolutionary distance of 0.1. We study four representative likelihood methods (fastDNAml, MrBayes, PAUP*-ML, TREE-PUZZLE) that use either Maximum Likelihood (ML) or Bayesian inference to find the optimal tree. PAUP*-ML denotes using the ML procedure in PAUP*4.0 [26]. NJ

Proceedings of the Third IEEE Symposium on BioInformatics and BioEngineering (BIBE'03)
0-7695-1907-5/03 $17.00 © 2003 IEEE

IEEE COMPUTER SOCIETY

is also used to provide a baseline comparison. For all experiments, MrBayes (a Bayesian inference approach) clearly outperformed its competitors.

The rest of this paper is organized as follows. Section II provides background information on using likelihood in phylogenetic analysis. Our experimental methodology is presented in Section III. Results are presented in Section IV. Finally, Section V provides concluding remarks.

II. Background

The principle of likelihood suggests that the explanation that makes the observed outcome the most likely occurrence is the one to be preferred. Formally, given data D and a model M, the likelihood of that data is given by $L = Pr(D|M)$, which is the probability of obtaining D given M. In the context of phylogenetics, D is the sequences of interest, and M is a phylogenetic tree. Below, we describe two approaches that use the principle of likelihood—Maximum Likelihood (ML) and Bayesian inference.

A. Maximum Likelihood

The Maximum Likelihood (ML) estimate of a phylogeny is the tree for which the observed data are most probable. Consider the data as aligned DNA sequences for n taxa. ML reconstruction consists of two tasks. The first task involves finding edge lengths that maximize the likelihood given a particular topology. Techniques to estimate branch lengths are based on iterative methods such as Expectation Maximization (EM) [2] or Newton-Raphson optimization [17]. Each iteration of these methods requires a large amount of computation. For a given tree, we must consider all possible nucleotides (A, G, C, T) at each interior node. The number of nucleotide combinations to examine for an n-taxa tree is $4^{(n-2)}$, since there are $n-2$ interior nodes.

The second task is to find a tree topology that maximizes the likelihood. Naïve, exhaustive search of the tree space is infeasible, since the number of bifurcating unrooted trees to be explored for n taxa is $\frac{(2n-5)!}{2^{n-3}(n-3)!}$. Furthermore, finding the best tree is hampered by the costly procedure of estimating edge lengths for different trees. Instead, the most likely tree is constructed in a greedy fashion by using a stepwise-addition algorithm. An initial tree is created by starting with three randomly chosen taxa. There is only one possible topology for a set of three taxa. A fourth randomly chosen taxon is attached, in turn, to each branch of the initial tree. The tree with the highest score is used as the starting tree for adding a fifth randomly chosen taxon. The process continues until all n taxa are added to the tree.

B. Bayesian inference

The Bayesian approach to phylogenetics builds upon a likelihood foundation. It is based on a quantity called the *posterior probability* of a tree. Bayes' theorem

$$Pr(Tree|Data) = \frac{Pr(Data|Tree) \times Pr(Tree)}{Pr(Data)}$$

is used to combine the *prior probability* of a phylogeny ($Pr(Tree)$) with the *likelihood* ($Pr(Data|Tree)$) to produce a *posterior probability* distribution on trees ($Pr(Tree|Data)$). The posterior probability represents the probability that the tree is correct. Note that, unlike likelihood scores, posterior probabilities sum to 1. Inferences about the history of the group are then based on the posterior probability of trees. For example, the tree with the highest posterior probability might be chosen as the best estimate of phylogeny. Typically all trees are considered a priori equally probable, and likelihood is calculated using a substitution model of evolution.

Computing the posterior probability involves a summation over all trees, and, for each tree, integration over all possible combinations of branch length and substitution model parameter values. A number of numerical methods are available to allow the posterior probability to be approximated. The most common is Markov Chain Monte Carlo (MCMC). For the phylogeny problem, the MCMC algorithm involves two steps. First, a new tree is proposed by stochastically perturbing the current tree. Afterwards, the tree is either accepted or rejected with a probability described by Metropolis *et al.* [13] and Hastings [5]. If the new tree is accepted, then it is subjected to more perturbations. For a properly constructed and adequately run Markov chain, the proportion of the time that any tree is visited is a valid approximation of the posterior probability of that tree.

C. ML versus Bayesian inference

Bayesian analysis of phylogenies is similar to ML in that the user postulates a model of evolution and the program searches for the best tree(s) that are consistent with both the model and the data. Both attempt to estimate a conditional probability density function: $Pr(Data|Tree)$ for ML and $Pr(Tree|Data)$ for Bayesian methods. The salient difference between the two methods is that Bayesian analysis requires an additional parameter, the prior probability, that must be fixed in advance. Since the prior probability is unknown, using any fixed guess introduces a potential source of error. Another difference that is often argued is that ML seeks the single most likely tree, whereas Bayesian analysis searches for the best set of trees leading to a construction of a consensus tree. However, such a difference is more a matter of traditional usage than one of algorithmic

IEEE COMPUTER SOCIETY

design. For example, the ML approach does not prohibit the user from keeping the top 10 trees and computing their consensus.

III. Experimental Methodology

We use a simulation study in order to evaluate the likelihood-based methods under consideration. In a simulated environment, the true (or model) tree is artificially generated and forms the basis for comparison among phylogenetic methods are compared. (With real data, the true tree is typically unknown). The simulation process starts with evolving a single DNA sequence (placed at the root of the model tree) under some assumed stochastic model of evolution to a set of sequences (located at the leaves of the tree). These sequences are given to the phylogenetic reconstruction methods, which infer a tree based on the given sequences. The inferred trees are then compared against the model tree for topological accuracy. The above process is repeated many times to obtain a statistically significant test of the performance of the methods under these conditions.

A. Model Trees

We used random birth-death trees as model trees for our experiments. The birth-death trees were generated using the program `r8s` [23], using a backward Yule (coalescent) process with waiting times taken from an exponential distribution. An edge e in a tree generated by this process has length λ_e, a value that indicates the expected number of times a random site changes on the edge. Trees produced by this process are ultrametric, meaning that all root-to-leaf paths have the same length. A random site is expected to change once on any root-to-leaf path; that is the (evolutionary) height of the tree is 1.

In our simulation study, we scale the edge lengths of the trees to give different heights (or evolutionary distances). To scale the length of the tree to height h, we multiply each edge of the tree by h. We used heights of 0.1, 0.3, and 1.0 in our study. We also modify the edge lengths of birth-death trees to deviate the tree away from the assumption that sites evolve under a strong molecular clock. To deviate a tree away from ultrametricity with deviation c we proceed as follows. For each edge $e \in T$, choose a number, x, uniformly from $[-\lg c, \lg c]$. Afterwards, multiply the branch length λ_e by e^x. Given deviation c, the expected deviation will be $(c-1/c)/2\ln c$. In our experiments, the deviation factor, c, was set to 4. Hence, the expected value each branch length is multiplied by is 1.36.

B. DNA Sequence Evolution

Once model trees are constructed, sequences are generated based on a model of evolution. The software used for sequence evolution is `Seq-Gen` [19]. We generated sequences under the Kimura 2-parameter (K2P) [10] + Gamma [28] model of DNA sequence evolution, one of the standard models for studying the performance of phylogenetic reconstruction methods. Under K2P, each site evolves down the tree under the Markov assumption. In this model, nucleotide substitutions are categorized based on the substitutions of pyrimidines (C and T) and purines (A and G). A *transition* substitutes a purine for a purine or a pyrimidine for a pyrimidine, whereas a *transversion* is a substitution of a purine for a pyrimidine or vice versa. The probability of a given nucleotide substitution depends on the edge and upon the type of substitution. We set the transition/transversion ratio, κ, to 2 (the standard setting). For rate heterogeneity, different rates were assigned to different sites according to the Gamma distribution with parameter α set to 1 (the standard setting).

C. Measures of Accuracy

We use the Robinson-Foulds (RF) distance [21] to measure the error between trees. For every edge e in a leaf-labeled tree, T defines a bipartition π_e on the leaves (induced by the deletion of e). The tree T is uniquely characterized by the set $C(T) = \{\pi_e : e \in E(T)\}$, where $E(T)$ is the set of all internal edges of T. If T is a model tree and T' is the tree obtained by a phylogenetic reconstruction method, then the error in the topology can be calculated as follows:

- False Positives (FP): $C(T') - C(T)$.
- False Negatives (FN): $C(T) - C(T')$.

The RF distance is the average of the number of False Positives and False Negatives. Our figures plot the *RF rates*, which are obtained by normalizing the RF distance by the number of internal edges. Thus, the RF rate varies between 0% and 100%.

D. Phylogenetic Reconstruction Methods

We considered five phylogenetic reconstruction methods in our simulation study. Three of the methods (fastDNAml, TREE-PUZZLE, and PAUP*-ML) use ML to construct the inferred tree. MrBayes is the sole algorithm that uses a Bayesian approach. NJ was used to provide a baseline comparison. The software package PAUP*4.0 was used to construct ML and NJ trees, which we denote by PAUP*-ML and NJ, respectively.

1) fastDNAml: fastDNAml [17] uses the stepwise-addition algorithm (described in Section II-A) to build a tree consisting of n-taxa. The best tree resulting from each step is subjected to local rearrangements, which explore the search space for a more likely tree. After all n taxa are added to the tree, a global rearrangement stage may be invoked.

IEEE COMPUTER SOCIETY

2) MrBayes: MrBayes [6] is a Bayesian inference method that incorporates both Markov Chain Monte Carlo (MCMC) (described in Section II-B) and Metropolis-Coupled Markov Chain Monte Carlo (MC^3) analysis [4]. MC^3 can be visualized as a set of independent searches that occasionally exchange information, which may allow a search to escape an area of local optima. For both types of analysis, the final output is a set of trees that the program has repeatedly visited. A majority rule consensus tree may be built from the output trees.

3) TREE-PUZZLE: TREE-PUZZLE [24] is a quartet-puzzling (QP) algorithm. Let S be the set of sequences labeling the leaves of an evolutionary tree T. A *quartet* of S is a set of four sequences taken from S, and a *quartet topology* is an evolutionary tree on four sequences. The QP algorithm consists of three steps. In the *ML step*, all $\binom{n}{4}$ quartet ML trees are reconstructed to find the most likely relationship for each set of four out of n sequences. During the *puzzling step*, these quartet trees are composed into an intermediate tree adding sequences one by one. The result of this step is highly dependent on the order of sequences. As a result, many intermediate trees from different input orders are constructed. From these intermediate trees a consensus tree (typically using a majority rule) is built in the *consensus step*.

4) PAUP-ML :* Similarly to fastDNAml, PAUP*-ML uses a stepwise-addition strategy to build the final tree. Local rearrangements occur after each taxon has been added to find a better tree. Once the final taxon has been added, a series of global rearrangements occur to find the final tree.

5) Neighbor-Joining: Neighbor-Joining [22] is one of the most popular distance-based methods. NJ use a distance matrix to compute the resulting tree. For every pair of taxa, it determines a score based on the distance matrix. The algorithm joins the pair with the minimum score, making a subtree whose root replaces the two chosen taxa in the matrix. Distances are recalculated based on this new node, and the "joining" continues until three nodes remain. These nodes are joined to form an unrooted binary tree. Appealing features of NJ are its simplicity and speed.

IV. Experimental Results

Our simulation study explores the behavior of the phylogenetic methods under a variety of different conditions. We generated 10 model trees for each tree size (20, 40, and 60 taxa) and scaled their heights by 0.1, 0.3, and 1.0 to vary the evolutionary distance. Afterwards, DNA sequences of lengths 100, 200, 400, 800, and 1600 were created from the scaled trees. All programs were run with the recommended default settings on Pentium 4 computers. Since MrBayes incorporates MC^3 analysis, we ran MrBayes with 2 chains, denoted MrBayes-CH2. Section IV-D considers the performance of MrBayes when the number of chains varies. For MrBayes, we used 60,000 generations, the current tree was saved every 100 generations, and consensus analysis ignored the first 120 trees (the burn-in value).

A. Speed

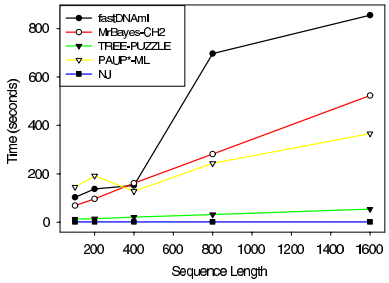
Figure 1 presents the execution time required by the different phylogenetic methods. NJ is fast and requires less than 1 second to output a final tree. The likelihood approaches, on the other hand, require large amounts of time to complete a phylogenetic analysis. For 20 taxa, TREE-PUZZLE is the fastest likelihood-based algorithm, but the algorithm is unable to complete an analysis for either 40 or 60 taxa. (A three-day limit was placed on an experimental run.) The maximum likelihood calculation of $\binom{n}{4}$ quartet trees is the source of TREE-PUZZLE's inability to produce an answer within a reasonable amount of time.

MrBayes is the only likelihood-based method that consistently produced a tree on all of our datasets. Its running time scales linearly with the sequence length. Neither fastDNAml nor PAUP*-ML terminated for 60 taxa. Instead, these ML methods cycled infinitely among equally likely trees. With increases in both the number of taxa and sequence length, the likelihood score approaches negative infinity, which cannot be represented precisely in computer hardware. Bayesian approaches avoid infinite loops since the user must specify the number of generations an analysis will execute. (PAUP*-ML provides an option to place a time limit on the heuristic search, but we did not use it.) For 40 taxa, the ML methods require 1–3 hours of computation time. MrBayes, on the other hand, requires less than 30 minutes of execution time on all datasets. MrBayes is quite fast for a likelihood-based approach. However, if compared to other methods (i.e., NJ, Weighbor, and Greedy MP), its slow nature becomes apparent [16].

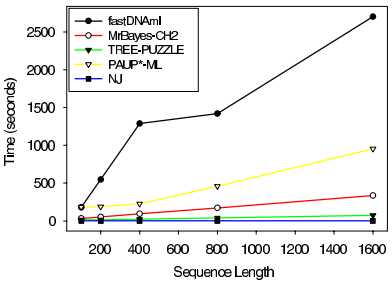
B. Accuracy

Figure 2 shows the accuracy of the phylogenetic methods as a function of the sequence length. Although NJ is computationally fast, it is the worst performer in terms of accuracy. NJ does outperform TREE-PUZZLE in Figure 2(a). All methods improve their RF rates as the sequence length is increased. Our results are consistent with those of Nakhleh *et al.* in that NJ's ability to produce a topologically correct tree degrades rapidly with an increase in evolutionary distance [16]. MrBayes and PAUP*-ML are less susceptible to changes in accuracy as the evolutionary distance increases. When PAUP*-ML terminated, the tree it produced was quite accurate. fastDNAml, on the other hand, finds it more

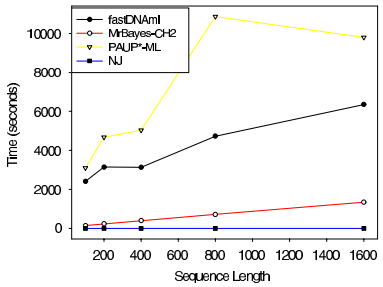
IEEE COMPUTER SOCIETY



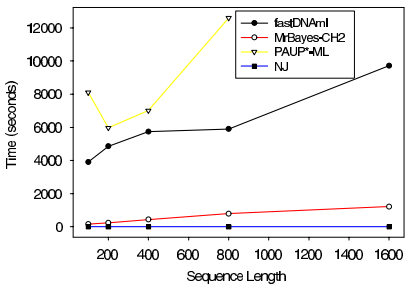
(a) Taxa = 20, Height = 0.3



(b) Taxa = 20, Height = 1.0

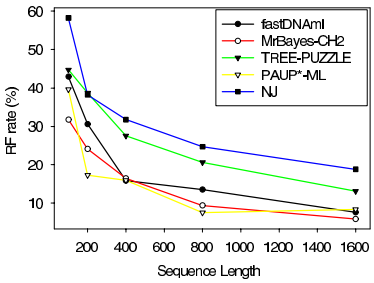


(c) Taxa = 40, Height = 0.3

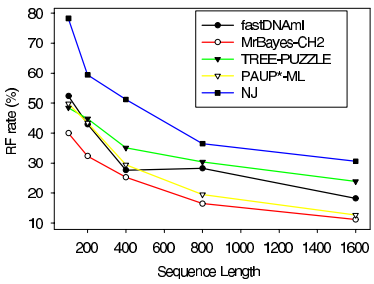


(d) Taxa = 40, Height = 1.0

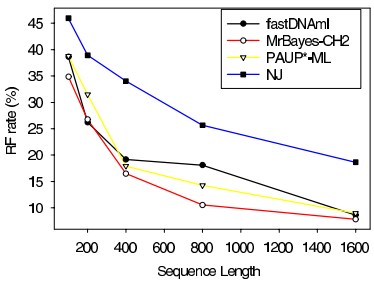
Fig. 1. Execution time of the phylogenetic methods as a function of the sequence length for various values of taxa and height.



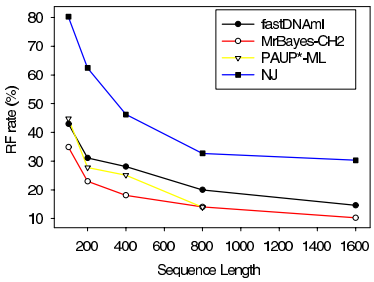
(a) Taxa = 20, Height = 0.3



(b) Taxa = 20, Height = 1.0



(c) Taxa = 40, Height = 0.3



(d) Taxa = 40, Height = 1.0

Fig. 2. Average RF rate of the phylogenetic methods as a function of the sequence length for various values of taxa and height.

Proceedings of the Third IEEE Symposium on BioInformatics and BioEngineering (BIBE'03)
0-7695-1907-5/03 $17.00 © 2003 IEEE

IEEE COMPUTER SOCIETY

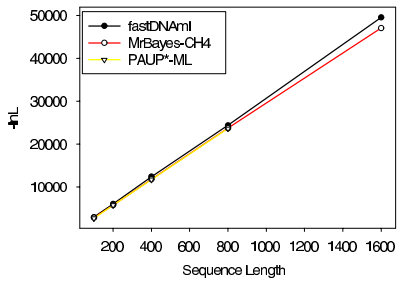


Fig. 3. Likelihood scores of fastDNAml, MrBayes-CH2, and PAUP*-ML as a function of sequence length for 40 taxa and a height of 1.0.

difficult to produce a topologically accurate solution on trees with large evolutionary distances.

At 60,000 generations, MrBayes converges to a stable posterior estimate. However, the number of generations affects the overall running time. We considered the accuracy of MrBayes as the number of generations varies. A shorter number of generations (20,000 and 40,000) resulted in a serious drop in topological accuracy. Increases in the number of generations (100,000) resulted in relatively small changes in accuracy (not shown).

C. Likelihood scores

The likelihood, L, of a tree is often a very small number. So, likelihoods are expressed as natural logarithms and referred to as log likelihoods, $\ln L$. Since the log likelihood of a tree is typically negative, our graphs plot positive log likelihoods or $-\ln L$. Consequently, in our graphs, the best likelihood is the one with the lowest score.

Figure 3 plots the likelihood scores of fastDNAml, MrBayes-CH2, and PAUP*-ML as a function of sequence length. The likelihood scores increase linearly with sequence length. Similar behavior was observed with an increase in the number of taxa (not shown). It is widely believed that higher likelihood scores correlates with an increase in topological accuracy. Yet, our results do not support this hypothesis. Although the likelihood scores returned by the phylogenetic methods are practically indistinguishable, the error rates are not. Once a method reaches its top trees, their likelihood values are so similar that selecting a tree essentially becomes a random choice.

D. MrBayes: A Closer Examination

Our results clearly establish MrBayes as the best likelihood-based method, so we take a closer look at its performance. MrBayes is capable of both Markov Chain Monte Carlo (MCMC) and Metropolis-Coupled Markov Chain Monte Carlo (MC^3) analysis. Proponents of the MC^3 method claim that multiple chains enable the search to avoid being trapped in a suboptimal region. We study the performance of MrBayes when 1, 2, and 4 chains are used.

Figure 4 plots the execution time of MrBayes for different numbers of chains as a function of the sequence length. Clearly, using multiple chains requires additional execution time. The increase in running time is due to chains swapping with each other to explore different parts of the tree space. An MCMC (single chain) analysis avoids such overhead. However, are multiple chains worth the extra computational effort? In many cases, the answer is no (see Figure 5). Running MrBayes with a single chain produced fairly good trees for low evolutionary rates (0.1, 0.3). As the evolutionary rates increase, multiple chains provide better solutions.

The results above show that MrBayes achieves consistent accuracy if two chains are used. Figure 6 plots the average RF rates of MrBayes-CH2 and NJ on trees of various heights. In this figure, MrBayes-0.1, MrBayes-0.3, and MrBayes-1.0 are the plots of MrBayes-CH2 for trees of height 0.1, 0.3, and 1.0, respectively. The accuracy of NJ degrades dramatically as the evolutionary distance is increased. MrBayes, on the other hand, seems to overcome this problem. On trees of large height, MrBayes-1.0 approaches the topological accuracy achieved on trees of lower height.

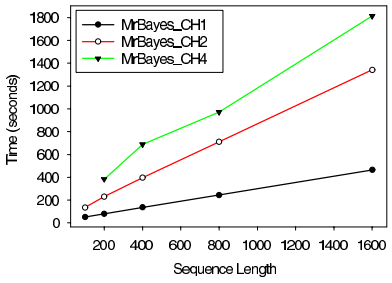
V. Concluding Remarks

This is the first study (to the best of our knowledge) to experimentally study the performance of likelihood methods as the number of taxa, sequence length, and evolutionary distance vary. Our results clearly demonstrate that MrBayes is the best algorithm of the methods we studied. The ML methods (fastDNAml, PAUP*-ML, and TREE-PUZZLE) have difficulty producing an answer on a consistent basis. (When PAUP*-ML converged to a solution, the inferred tree was quite good.)

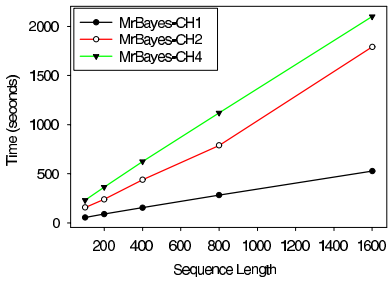
We recommend running MrBayes with two chains for phylogenetic analysis. MC^3 analysis seems to be best suited for large evolutionary distances. Although MrBayes runs relatively quickly, its running time is still a limiting factor. One cannot expect to use MrBayes, in its current form, to solve large datasets (of several thousand taxa). Although MrBayes produced the best trees, its likelihood score was indistinguishable from the scores returned by the other methods. For the top trees of a heuristic search, likelihood scores do not correlate well to topological accuracy. These results do not imply that such scores are unimportant—just that the correlation, which clearly exists over the tree space as a whole, is lost in regions of high likelihood.

There are many directions for future work. Combining disk-covering methods (DCMs) [8], [9] with Bayesian inference is one approach to speeding up a likelihood

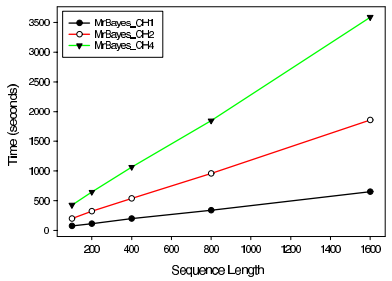
IEEE COMPUTER SOCIETY



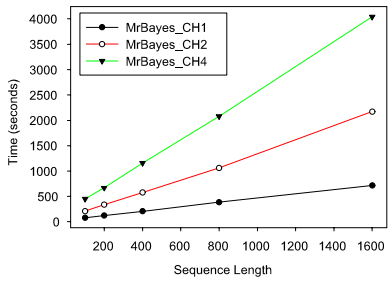
(a) Taxa = 40, Height = 0.3



(b) Taxa = 40, Height = 1.0

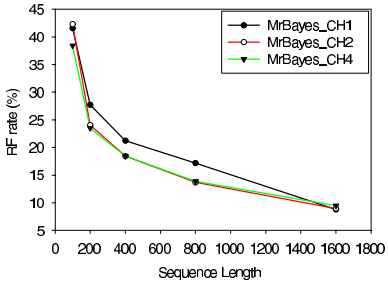


(c) Taxa = 60, Height = 0.3

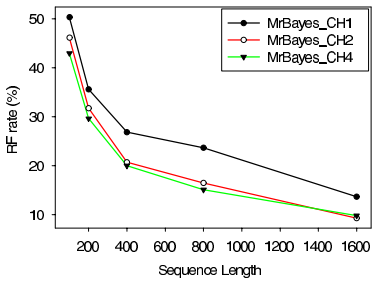


(d) Taxa = 60, Height = 1.0

Fig. 4. Execution time on MrBayes as a function of the sequence length for various values of taxa and height.

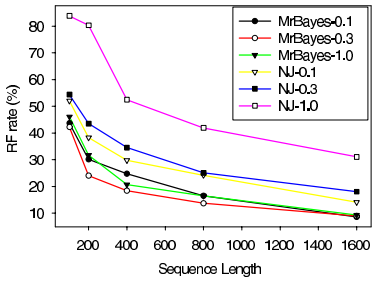


(a) Taxa = 60, Height = 0.3

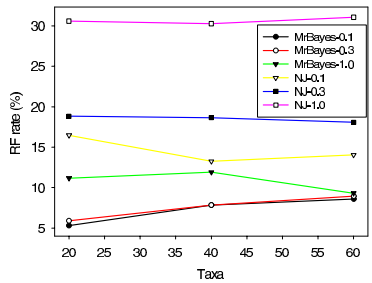


(b) Taxa = 60, Height = 1.0

Fig. 5. Average RF rate for MrBayes as a function of the sequence length for 60 taxa and heights 0.3 and 1.0.



(a) Taxa = 60



(b) Sequence Length = 1600

Fig. 6. Average RF rate for MrBayes and NJ as a function of (a) sequence length and (b) number of taxa. MrBayes-0.1, MrBayes-0.3, and MrBayes-1.0 are the plots of MrBayes-CH2 for trees of height 0.1, 0.3, and 1.0, respectively.

IEEE COMPUTER SOCIETY

calculation. The question of whether likelihood scores correlates with topological accuracy is an interesting one that needs further exploration.

VI. Acknowledgements

During the course of this research, Williams was supported by a Alfred P. Sloan Foundation Postdoctoral Fellowship in Computational Molecular Biology (Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG03-02ER63426), and Moret was supported by the National Science Foundation under grants ACI 00-81404, EIA 01-13095, EIA 01-21377, and DEB 01-20709.

References

[1] D. Bader, B. M. Moret, and L. Vawter. Industrial applications of high-performance computing for phylogeny reconstruction. In H. Siegel, editor, *Proceedings of SPIE Commercial Applications for High-Performance Computing*, volume 4528, pages 159–168, Denver, CO, Aug. 2001. SPIE.

[2] J. Felsenstein. Evolutionary trees from dna sequences: a maximum likelihood approach. *J. Mol. Evol.*, 17:368–376, 1981.

[3] N. Friedman, M. Ninio, I. Pe'er, and T. Pupko. A structural EM algorithm for phylogenetic inference. *J. Comp. Biol.*, 9:331–353, 2002.

[4] C. J. Geyer. Markov chain Monte Carlo maximum likelihood. In Keramidas, editor, *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pages 156–163. Interface Foundation, 1991.

[5] W. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[6] J. P. Huelsenbeck and F. Ronquist. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8):754–755, 2001.

[7] J. P. Huelsenbeck, F. Ronquist, R. Nielsen, and J. P. Bollback. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, 294:2310–2314, 2001.

[8] D. Huson, S. Nettles, and T. Warnow. Disk-covering, a fast-converging method for phylogenetic tree reconstruction. *Comput. Biol.*, 6:369–386, 1999.

[9] D. Huson, L. Vawter, and T. Warnow. Solving large scale phylogenetic problems using DCM2. In *ISMB99*, pages 118–129, 1999.

[10] M. Kimura. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, 16:111–120, 1980.

[11] B. Larget and D. L. Simon. Markov chain monte carlo algorithms for the bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.*, 16:750–759, 1999.

[12] P. O. Lewis. A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. *Mol. Biol. Evol.*, 16:277–283, 1998.

[13] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[14] M. L. Metzker, D. P. Mindell, X.-M. Liu, R. G. Ptak, R. A. Gibbs, and D. M. Hillis. Molecular evidence of HIV-1 transmission in a criminal case. *PNAS*, 99(2):14292–14297, 2002.

[15] B. M. Moret, U. Roshan, and T. Warnow. Sequence length requirements for phylogenetic methods. In *Proc. 2nd Int'l Workshop on Algorithms in Bioinformatics (WABI'02)*, volume 2452 of *Lecture Notes in Computer Science*, pages 343–356, 2002.

[16] L. Nakhleh, B. Moret, U. Roshan, K. S. John, and T. Warnow. The accuracy of fast phylogenetic methods for large datasets. In *Proc. 7th Pacific Symp. on Biocomputing (PSB 2002)*, pages 211–222. World Scientific Pub. (2002), 2002.

[17] G. J. Olson, H. Matsuda, R. Hagstrom, and R. Overbeek. fastDNAml: A tool for construction of phylogenetic trees of dna sequences using maximum likelihood. *Comput. Appl. Biosci.*, 10:41–48, 1994.

[18] R. D. M. Page and E. C. Holmes. *Molecular Evolution: A Phylogenetic Approach*. Blackwell Science Inc., 1998.

[19] A. Rambaut and N. C. Grassly. Seq-gen: An application for the monte carlo simulation of dna sequence evolution along phylogenetic trees. *Comp. Appl. Biosci.*, 13:235–238, 1997.

[20] V. Ranwez and O. Gascuel. Quartet-based phylogenetic inference: Improvements and links. *Mol. Biol. Evol.*, 18:1103–1116, 2001.

[21] D. F. Robinson and L. R. Foulds. Comparison of phylogenetic trees. *Mathematical Biosciences*, 53:131–147, 1981.

[22] N. Saitou and M. Nei. The neighnbor-joining method: A new method for reconstructiong phylogenetic trees. *Mol. Biol. Evol.*, 4(406–425):1987, 1987.

[23] M. J. Sanderson. *r8s software package*. `Available from http://ginger.ucdavis.edu/r8s/`.

[24] H. Schmidt, K. Strimmer, M. Vingron, and A. von Haeseler. Tree-puzzle: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, 18:502–504, 2002.

[25] K. St. John, T. Warnow, B. M. Moret, and L. Vawter. Performance study of phylogenetic methods: (unweighted) quartet methods and neighbor-joining. *J. Algorithms*, 2003. to appear.

[26] D. Swofford. *PAUP*:Phylogenetic analysis using parsimony (and other methods)*. Sinauer Associates, Sunderland, MA, 1996. Version 4.0.

[27] T. L. Williams and B. M. Moret. An investigation of phylogenetic likelihood methods. Technical Report TR-CS-2002-33, The University of New Mexico, 2002.

[28] Z. Yang. Maximum likelihood estimation of phylogeny from dna sequences when substitution rates differ over sites. *Mol. Biol. Evol.*, 10:1396–1401, 1993.

IEEE COMPUTER SOCIETY